(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 October 2001 (04.10.2001)

PCT

(10) International Publication Number
WO 01/73134 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US01/09993

(22) International Filing Date: 28 March 2001 (28.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/192,700 28 March 2000 (28.03.2000) US

(71) Applicant (*for all designated States except US*): **THE GOVERNMENT OF THE UNITED STATES OF AMERICA, as represented by THE SECRETARY, DEPARTMENT OF HEALTH & HUMAN SERVICES, THE NATIONAL INSTITUTES OF HEALTH** [US/US]; Office Of Technology Transfer, 6011 Executive Boulevard, Suite #325, Rockville, MD 20852 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **WANG, Ena** [CN/US]; 4500 Gretna Street, Bethesda, MD 20814 (US). **MARINCOLA, Francesco, M.** [US/US]; 10110 Chapel Road, Potomac, MD 20854 (US). **MILLER, Lance, D.** [US/US]; 13109-E Millhaven Place, Germantown, MD 20874 (US).

(74) Agent: **HARDING, Tanya, M.**; Klarquist, Sparkman, Campbell, Leigh & Whinston, LLP, One World Trade Center, Suite 1600, 121 S.W. Salmon Street, Portland, OR 97204 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: GENE PROFILING ARRAYS

(57) **Abstract:** Ordered arrays of mixtures of nucleic acid molecules are provided, which mixtures reflect the expression profile of one or more specimens, such as different cells or tissues. In particular embodiments, complete mRNA mixtures from specimens are separately arrayed on a substrate. Specimens from which such mixtures of nucleic acid molecules are produced can be taken from any source, including animal, plant and/or microbial cells, and can be assembled in any collection desired. The collections can, for instance, include different cell types, different phenotypes, cells grown under different conditions, cells of different ages or developmental stages, and so forth. The nucleic acid arrays are provided in both macro- and microarray formats, and are suitable for gene profiling in which relative quantitative expression from a single source or multiple sources may be determined. Techniques are also disclosed for producing high-fidelity, amplified mixtures of nucleic acid molecules using a combination of anti-sense RNA amplification and template-switching synthesis. Amplified mixtures produced using this method can, for instance, be applied to the disclosed arrays. The disclosed arrays allow high throughput analysis of differential gene expression in a specimen (such as a tumor) or a variety of specimens (such as a variety of tumors), and are suitable for automated preparation and analysis.

WO 01/73134 A2

GENE PROFILING ARRAYS

FIELD

The present disclosure relates to methods and devices useful for analyzing the gene expression, particularly for comparing gene expression in a plurality of cells or tissues simultaneously. The disclosure also relates to the preparation of nucleic acid samples useful in such simultaneous analysis of gene expression.

BACKGROUND

Current microarray technology typically involves depositing nucleic acids on a solid platform in a set pattern, and hybridizing a solution of heterogeneous, labeled, potentially complementary nucleic acids to the nucleic acid targets. Microarray technology is used to detect mutations and polymorphisms, to compare gene expression profiles, and for genotyping, genetic mapping, and DNA sequence analysis, depending on the nucleic acids used as target and probe. For an overview of this technology, see Gerhold *et al.*, *TIBS* 24:168-173, 1999, and Epstein & Butow, *Current Opin. Biotech.* 11:36-41, 2000.

A specific example of a conventional microarray is a "cDNA microarray," on which samples of individual (usually known) cDNA molecules or fragments thereof are arrayed ("spotted") on a solid microarray substrate such as a chip, glass slide or supported membrane. Each addressable (capable of being reliably and consistently located and identified) spot on the array contains only one cDNA sequence, though there are many copies of the sequence in the spot. A cDNA microarray can be used to compare gene expression profiles from two tissues/cells by exposing the array to labeled nucleic acid from the different tissues or cell types. Differences in the hybridization signal intensity at a single microarray locus (which corresponds to a single arrayed cDNA sequence) are indicative of differences in the expression of the corresponding message in the tested tissues.

Techniques currently used to prepare material for analysis of conventional cDNA microarrays require a relatively large quantity of RNA, either mRNA or total RNA, to prepare the labeled nucleic acid from the sample. Incyte, a company that prepares and runs GEM™ microarray analyses using researcher materials, requires about 100 µg of total RNA, or 600 ng of polyA RNA, to prepare enough probe for one hybridization to a standard microarray (see, for instance, documentation posted on the synteni.com web site). The Affymetrix microarray system requires about 3-5 µg of mRNA, or about 5-50 µg of total RNA (discussed in Gerhold *et al.*, *TIBS* 24:168-173, 1999). Duggan *et al.* (*Nat. Gen. Suppl.* 21:10-14, 1999), in reviewing the "clear limitation" of microarray technology requiring a "large amount of RNA" for each hybridization, discuss use of 50-200 µg of total, or 2-5 µg of polyA RNA. Research groups report using various amounts of starting material; Schena *et al.* (*Science* 270:467-470, 1995) fluorescently labeled 5 µg of mRNA; Chen *et al.* (*Genomics* 51:313-324, 1998) biotin-labeled 2 µg of mRNA; and Lockhart *et al.* (*Nat. Biotech.* 14:1675-1680, 1996) started with 1 µg of polyA RNA. Because so much starting material is

required, certain clinical samples, such as small biopsies or individual cells, are considered inadequate for production of a microarray probe.

Systems currently used to produce nucleic acids for microarray probes do not maintain proportionality of individual messages during amplification, and reproduction of full-length cDNA is not guaranteed (Carulli *et al.*, *J. Cell. Biochem. Suppl.* 30/31:286-296 at 290, 1998). Absence of proportionality is inherent in the use of the most common method of second strand cDNA synthesis, the Gulber-Hoffman method (*Gene* 25:263-269, 1983), in which second strand cDNA synthesis is primed from randomly nicked mRNA using RNase H, DNA polymerase I, and DNA ligase.

In addition, current microarray technology only permits the analysis of the expression of a collection of known (arrayed) gene sequences in a single target cell from which a heterogeneous pool (mixture) of nucleic acid molecules are isolated and labeled.

Traditional Northern blots and Northern dot blots are systems used to compare the entire expression profile of one or a few genes from multiple cells/tissues at the same time on the same substrate. In a Northern blot, RNA molecules (targets, typically mRNA) are extracted from a plurality of different samples (*e.g.*, different cells, tissues, or species) and "run out" on a gel to separate the nucleic acid molecules based at least in part on their molecular weights. The content of the resultant gel is then transferred to nitrocellulose membrane or another such substrate, and hybridized to a labeled nucleic acid sample containing a single sequence of interest (*e.g.*, corresponding to a gene for which expression data is desired). Northern dot blotting involves binding mRNA extracts from different samples to a nitrocellulose membrane or other suitable substrate by application through a "dot blot" or "slot blot" apparatus (for an example, see the "Bio-Dot Microfiltration Apparatus" produced by Bio-Rad Laboratories, Hercules, CA). This is similar to a Northern blot, except that there is no primary separation of the mRNA molecules in a gel.

Once mRNA extracts are bound to the membrane in the lanes corresponding to a gel (Northern) or as individual heterogeneous spots (Northern dot), the blot can be hybridized to a labeled nucleic acid sequence. Thus, the probing molecule is a labeled, known nucleic acid sequence that hybridizes to heterogeneous mixtures of nucleic acid targets on the surface of the substrate.

Northern (dot) blot techniques are cumbersome and tedious, requiring extensive handling of RNA, which is an inherently fragile molecule prone to degradation in the laboratory. In addition, both of these techniques require technician involvement at several stages, and do not lend themselves well to automation. These traditional techniques also require a large amount of starting material to provide an interpretable signal, and thus cannot be used to analyze certain specialized or low abundance cell or issue types (*e.g.*, fine needle aspirates or micro-dissection or experimental models studying embryonic tissue or small organisms).

There still exists a need for methods, and devices to use therewith, that provide simple, automatable, high throughput techniques for simultaneous analysis of gene expression in a plurality of cells or tissues and which do not require large amounts of starting materials. The current invention is directed to addressing this need.

SUMMARY

Devices and methods disclosed herein overcome several disadvantages of existing methods of gene expression analysis.

Using the arrays and methods described herein, thousands of different kinds of cell types and tissues can be analyzed for gene expression simultaneously. An expression profile can be determined for each gene product of interest. In addition, multiple genes can be simultaneously profiled using probes labeled with different fluorescent labels. Since these gene profiling cDNA library arrays are much more stable than mRNA arrays used for Northern blots, they can be widely applied to laboratory situations without requiring stringent experimental conditions. In addition, the cDNA molecules of the array are naturally antisense and therefore bind well with sense-strand probes.

Certain embodiments are assay methods useful for determining gene expression or for examining and measuring relative expression of a DNA sequence in a plurality of biological specimens. Such methods include providing an array of nucleic acid mixtures at addressable locations (*e.g.*, discrete locations such as spots) on a substrate. In particular embodiments, the nucleic acid mixtures include nucleic acid molecules in quantities that are substantially proportional to the quantities of the nucleic acid molecules in a specimen from which the nucleic acid molecules are obtained, and exposing the array to a probe. The probe may represent a gene product of interest, and is complementary to and specifically hybridizable to a target nucleic acid sequence. Such probes can be used for detecting one or more nucleic acid molecules on the array under conditions sufficient to produce binding of the probe to the one or more nucleic acid molecules in the arrayed mixtures of nucleic acid molecule. Optionally, the methods can also include detecting hybridization (binding) of the probe to one or more nucleic acid molecules immobilized on the array (if such hybridization occurs). Also, such methods can optionally include separating any unbound (unhybridized or non-specifically hybridized) probe from the array prior to detecting such binding. Detection can, in certain embodiments, include automated detection (*e.g.*, detection that is assisted by or carried out by a computer or system including a computer). Detection can also include detection of a binding pattern.

In specific embodiments, detecting binding or hybridization of the probe includes quantitatively detecting such binding to yield an amount of bound probe (hybridization). This amount can then be correlated with the expression levels of RNA molecules, and thus with a level of gene expression in the specimen that served as the source of the RNA molecules used to produce the mixture of nucleic acids on the array.

Examples of probes for use with arrays are nucleic acid molecules, for instance nucleic acid molecules having specific complementarity to a target RNA or RNA-derived molecule. Such probes can be single-stranded nucleic acid (*e.g.*, DNA) molecules. Probes can be made detectable, for

instance by the inclusion of a detectable tag, such as a fluorophore, a radioactive isotope, a ligand, a chemiluminescent agent, a metal sol, a metal colloid, or an enzyme.

Also provided are methods for examining the expression of more than one gene using the same array, by the sequential or simultaneous application of a plurality of different detectable probe molecules directed to (capable of hybridization with) different gene products. In such methods, especially those in which the different probes are applied to a single array simultaneously or without intervening stripping, it is beneficial to differently label the plurality of probes, for instance with fluorophores of different colors (*e.g.*, red and green). Different probes can be directed to different target molecules of interest, or to at least one control molecule (either a positive or negative control molecule) and at least one target molecule of interest. Specific examples of control molecules are housekeeping genes (and sequences derived from such housekeeping genes).

In certain examples, the nucleic acid mixtures of the array are stably associated with a surface of the substrate of the array, and can be arranged in regular or irregular patterns. The pattern is optimally "addressable" in that the position of each "spot" of nucleic acid mixture can be consistently and repeatedly correlated to the source specimen from which the mixture was derived.

Many types of specimens can be used as source material for the mixtures of nucleic acid arranged in the arrays. In certain embodiments, the specimens are selected from the group consisting of cells or tissues, for instance cells taken from animals, microbes, or plants. In specific embodiments, the animal cells can include human cells.

In certain examples of the disclosed methods, each mixture of nucleic acid molecules substantially proportionately reflects the expression level of substantially all expressed mRNA molecules of that specimen. In some embodiments, mixtures of nucleic acid molecules can be amplified, for instance by polymerase chain reaction prior to being detected by a probe or even prior to placement of the nucleic acid molecules on the array.

By way of example, one method for amplification includes isolating an RNA sample from a specimen; obtaining one or more RNA templates from a portion of the RNA sample; hybridizing the one or more templates with a first primer (*e.g.*, a primer that includes an antisense sequence of an RNA polymerase promote) to form a primed template; and synthesizing first strand cDNA from the primed template. A second primer (which includes a string of dG residues at the 3' end) is then hybridized to the first strand cDNA a to form a switched template, and this switched template is used to synthesize second strand cDNA, thereby generating full-length double stranded cDNA. Antisense RNA (aRNA) can be transcribed from the full-length double stranded cDNA; and amplified cDNA optionally reverse transcribed from the aRNA. Mixtures of nucleic acid molecules produced by this method are also encompassed, as are uses for such mixtures.

Also provided are gene profiling arrays, which include a plurality of mixtures of nucleic acid molecules, usually immobilized on a solid support (*e.g.*, glass, nitrocellulose, polyvinylidene fluoride, nylon, fiber, or combinations thereof) in an addressable pattern. In some embodiments of these arrays, each mixture of nucleic acid molecules proportionately reflects the expression levels of

mRNA molecules in a specimen from which the nucleic acid mixture was obtained. In certain arrays, the addressable pattern of mixtures of nucleic acid molecules is arranged in discrete spots, for instance arranged in rows and columns. In particular embodiments, the addressable pattern of such arrays can be arranged in a computer readable format, in which the spots are at addresses that are stored in or can be determined by an automated device that interprets hybridization signals (including their absence or intensity) at address of the array.

The different mixtures of nucleic acid molecules can be derived from a plurality of different specimens (such as tissues or cells derived from animals, plants or microbe). However, in certain specific embodiments samples of the same mixture of nucleic acid molecules (representing the same source specimen) will be applied to the same array. Alternatively, multiple samples of the same mixture(s) can be provided on the array with different mixtures from different specimens. Such duplicative applications can serve, for instance, to provide internal hybridization controls. Also, it is envisioned that different amounts of the samples may be applied to the substrate in forming the array, for instance to determine the optimal amount of mixture for hybridization experiments.

Arrays contemplated herein can contain, for instance, at least 10 different mixtures of nucleic acid molecules each located in a discrete spot, but may contain at least 30, at least 100, at least 1000, or more different mixtures in discrete spots. In particular embodiments the array is a microarray, for example in which spots on the array have a maximum dimension of about 1 millimeter.

Further embodiments are kits for determining relative expression of a DNA sequence of interest in a plurality of biological specimens (*e.g.*, tissues and/or cells from animals, plants and/or microbes), such kits including a gene profiling array as described herein, and instructions for using the array. These kits may further include one or more probes representing the DNA sequence of interest, and/or one or more probe standards (control probes), and/or one or more buffers. Probes included in these kits can optionally include a detectable tag or other label. In certain kits, the gene profiling array will include a microarray.

In further examples of these methods at least half of the mixtures of nucleic acid molecules on the array (either macro- or microarray) are from different specimens (*e.g.*, at least 10 different specimens or at least 100 different specimens on a single array). Specific embodiments of such methods are included, wherein at least one mixture of nucleic acids is derived from a specimen consisting of not more than 10 cells, and in certain embodiments the specimen consists of not more than one cell.

Particular methods are also disclosed wherein at least one nucleic acid mixture on the array is derived (*e.g.*, amplified) from a source RNA sample extracted from a source specimen, and wherein the source RNA sample consists of no more than about 1 μg of total RNA. In further specific embodiments, the source RNA sample consists of no more than about 0.75 μg of total RNA, no more than about 0.5 μg of total RNA, or no more than about 0.3 μg of total RNA.

Certain embodiments are based on the utilization of *in vitro* transcription to generate full length antisense amplified RNA (aRNA) with high fidelity. Because of the high-efficiency of the amplification, minimal amounts of total RNA can be amplified up to about 80,000-fold, generating pure aRNA without losing linearity. Thus-generated aRNA from different samples can be transcribed into antisense cDNA, and the resultant mixtures of cDNAs then printed onto arrays. Each spot on the array can represent a unique cDNA library pool (mixture) from a different specimen, which will often proportionately reflect the expression levels of each of the individual mRNAs in the source.

Certain disclosed embodiments also provide procedures that optimize amplification of low-abundance-RNA samples by combining anti-sense RNA (aRNA) amplification with template-switching synthesis. The fidelity of aRNA amplified from 1/10,000 to 1/100,000 of commonly used input RNA is comparable to expression profiles observed with conventional poly(A)-RNA (RNA that includes a poly-adenine tail) or total RNA-based arrays.

The foregoing and other features and advantages will become more apparent from the following detailed description of several embodiments, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a graphic representation of an assessment of random or labeling bias by hybridization of differentially labeled aRNA (3 µg) amplified from the same melanoma cell line (A375, ATCC, Rockville, MD) to 2008-gene OncoChip (top panel). Cy3 (green) *vs.* Cy5 (red) signal intensity for each spot was highly correlated (R^2 = 0.99). A similar scatter plot (bottom panel) compares aRNA from a melanoma (A375) *vs.* a lymphoid (ML-1, ATCC, Rockville, MD) cell line (labeled with Cy3 and Cy5 respectively); these two cell lines exhibit substantial differences in gene expression (R^2 = 0.28).

Figure 2A is a series of bar graphs showing grading of outlier reproducibility in mRNA, total RNA, and aRNA hybridizations. Mutually exclusive confidence groups of outliers (4, 3, 2 rec and 2 rep match) were defined by four consecutive total RNA-based (T-RNA) control hybridizations (see Example 1). Percentage of the genes belonging to each confidence group identified as outliers in experimental conditions are shown as bars. RNA concentrations in the labels refer to starting amount of source total RNA (see figure legend).

Figure 2B is a high-stringency hierarchical cluster diagram of differentially expressed genes (outliers) in mRNA, total RNA (T-RNA) and aRNA array hybridizations that encompasses all four confidence groups. Columns designate single array hybridizations: targets from melanoma cell lines are Cy3 (green) biased except for total RNA in which targets were reciprocally labeled (T-RNA-R). Numbers in parenthesis refer to amount of source total RNA from which aRNA was amplified. 2* refers to aRNA obtained after two rounds of amplification. Rows designate single genes (arrayed on the microarray described in Example 1). Green and red cells reflect genes expressed at higher levels in A375 (melanoma) and ML-1 (lymphoid) cells, respectively. Black cells indicate genes with

approximately equivalent expression levels and gray cells indicate missing or filter-excluded data. The magnitude of the log-transformed ratio is reflected by the degree of color saturation (see color scale at the bottom of the figure). The 251 genes with expression ratios of 3-fold or greater in at least five hybridizations are shown.

Figure 3A is a low-stringency cluster diagram of reproducible and anomalous (discordant) outliers. The 817 genes with 3-fold or greater expression ratios in at least one hybridization are shown. The blue bar to the right of the cluster diagram parallels a sub-cluster of anomalous outliers with minimal reproducibility, which were characterized by low signal intensity. Gray cells depict genes with missing data or signal intensities below 150 units in one or both channels. (Signal intensities are measured on a scale from 1 to 65,536 units.)

Figure 3B is a bar graph representing the measurement of experimental outliers discordant from the "true outliers" determined by the control total RNA hybridizations, presented as percentage of the total number of genes in the array.

Figure 4 is a schematic outline showing construction and probing (with a labeled (*) probe) of a gene profiling array.

Figure 5 is a schematic outline showing construction and probing (with a labeled (*) probe) of a gene profiling array wherein two signal intensities are detected.

Figure 6 is a schematic representation of a disclosed system for producing substantial amounts of high-fidelity full-length nucleotides, in the form of aRNA or cDNA produced from that aRNA, from a very small amount (as little as 0.5 μg) of starting total RNA.

SEQUENCE LISTING

The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand.

SEQ ID NO: 1 shows an oligo-dT primer used to prepare aRNA from total RNA in the disclosed high-fidelity amplification system.

SEQ ID NOs: 2 and 3 show template switch primers used in the disclosed system for high-fidelity amplification of mRNA.

DETAILED DESCRIPTION

I. Abbreviations and Terms

A. Abbreviations

aRNA: antisense messenger RNA (also asRNA)

cDNA: complementary DNA

DNA: deoxyribonucleic acid

EST: expressed sequence tag

PNA: peptide nucleic acid

B. Terms

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes VII*, published by Oxford University Press, 2000 (ISBN 0-19-899276-X); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

In order to facilitate review of the various embodiments, the following explanation of terms is provided:

Addressable: capable of being reliably and consistently located and identified, as in an addressable location on an array.

Antisense RNA (aRNA): A molecule of RNA complementary to a sense (encoding) nucleic acid molecule. Often, aRNA is constructed by transcribing antisense strand RNA from a cDNA molecule.

Array: An arrangement of molecules, particularly biological macromolecules (such as polypeptides or nucleic acids) in addressable locations on a substrate. The array may be regular (arranged in uniform rows and columns, for instance) or irregular. The number of addressable locations on the array can vary, for example from a few (such as three) to more than 50, 100, 200, 500, 1000, 10,000, or more. A "microarray" is an array that is miniaturized so as to require microscopic examination for evaluation.

Within an array, each arrayed molecule is addressable, in that its location can be reliably and consistently determined within the at least two dimensions of the array surface. Thus, in ordered arrays the location of each molecule sample is assigned to the sample at the time when it is spotted onto the array surface, and a key may be provided in order to correlate each location with the appropriate target. Often, ordered arrays are arranged in a symmetrical grid pattern, but samples could be arranged in other patterns (*e.g.*, in radially distributed lines, spiral lines, or ordered clusters). Addressable arrays are computer readable, in that a computer can be programmed to correlate a particular address on the array with information (such as hybridization or binding data, including for instance signal intensity). In some examples of computer readable formats, the individual "spots" on

the array surface will be arranged regularly, for instance in a Cartesian grid pattern, that can be correlated to address information by a computer.

The sample application "spot" on an array may assume many different shapes. Thus, though the term "spot" is used throughout, it refers generally to a localized deposit of nucleic acid pool (*e.g.*, a pool of nucleic acid molecules that reflects the expression level of mRNA in a cell or tissue sample, also referred to as a mixture of nucleic acids or nucleic acid molecules), and is not limited to a round or substantially round region. For instance, substantially square regions of mixture application can be used with arrays encompassed herein, as can be regions that are substantially rectangular (such as a slot blot-type application), or triangular, oval, or irregular. The shape of the array substrate itself is also immaterial, though it is usually substantially flat and may be rectangular or square in general shape.

In certain example arrays, each mixture of nucleic acid molecules will be spotted onto the array twice to provide internal controls.

Binding or interaction: An association between two substances or molecules. The arrays are used to detect binding of a labeled nucleic acid molecule (termed a "probe" herein) to an immobilized nucleic acid molecule in one or more mixtures of nucleic acid molecules of the array. A probe "binds" to a nucleic acid molecule in a spot on an array of this invention if, after incubation of the probe (usually in solution or suspension) with or on the array for a period of time (usually 5 minutes or more, for instance 10 minutes, 20 minutes, 30 minutes, 60 minutes, 90 minutes, 120 minutes or more), a detectable amount of the probe associates with a nucleic acid mixture of the array to such an extent that it is not removed by being washed with a relatively low stringency buffer (*e.g.*, higher salt (such as 3 x SSC or higher), room temperature washes). Washing can be carried out, for instance, at room temperature, but other temperatures (either higher or lower) can also be used. Probes will bind nucleic acid molecules within different immobilized nucleic acid of mixtures to different extents, and the term "bind" encompasses both relatively weak and relatively strong interactions. Thus, some binding will persist after the array is washed in a more stringent buffer (*e.g.*, lower salt (such as about 0.5 to about 1.5 x SSC), 55-65° C washes).

Where the probe molecule is a nucleic acid, binding of the probe molecule to a target can be discussed in terms of the specific complementarity between the probe molecule and the target nucleic acid.

The term "binding characteristics of an array for a particular probe" refers to the specific binding pattern that forms between the probe and the array after excess (unbound or not specifically bound) probe is washed away. This pattern (which may contain no positive signals, some or all positive signals, and will likely have signals of differing intensity) conveys information about the binding affinity of that probe for molecules within the spots of the array, and can be de-coded by reference to the key of the array (which lists the addresses of the spots on the array surface). The relative intensity of the binding signals from individual pool locations (spots) is indicative of the relative expression level of the nucleic acid that corresponds to the probe (at least to the extent that

the nucleic acid mixtures have been generated by a method that maintains the proportionality of each expression unit in the source material). Quantification of the binding pattern of an array/probe combination can be carried out using any of several existing techniques, including scanning the signals into a computer for calculation of relative density of each spot.

cDNA: A DNA molecule lacking internal, non-coding segments (introns) and regulatory sequences which determine transcription. cDNA may be synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

DNA: DNA is a long chain polymer that contains the genetic material of most living organisms (the genes of some viruses are made of ribonucleic acid (RNA)). The repeating units in DNA polymers are four different nucleotides, each of which includes one of the four bases (adenine, guanine, cytosine and thymine) bound to a deoxyribose sugar to which a phosphate group is attached. Triplets of nucleotides (referred to as codons) code for each amino acid in a polypeptide, or for a stop signal. The term "codon" is also used for the corresponding (and complementary) sequences of three nucleotides in the mRNA into which the DNA sequence is transcribed.

EST (Expressed Sequence Tag): A partial DNA or cDNA sequence, typically of between 500 and 2000 sequential nucleotides, obtained from a genomic or cDNA library, prepared from a selected cell, cell type, tissue or tissue type, organ or organism, which corresponds to an mRNA of a gene found in that library. An EST is generally a DNA molecule sequenced from and shorter than the cDNA from which it is obtained.

Fluorophore: A chemical compound, which when excited by exposure to a particular wavelength of light, emits light (*i.e.*, fluoresces), for example at a different wavelength. Fluorophores can be described in terms of their emission profile, or "color." Green fluorophores, for example Cy3, FITC, and Oregon Green, are characterized by their emission at wavelengths generally in the range of 515-540 λ. Red fluorophores, for example Texas Red, Cy5 and tetramethylrhodamine, are characterized by their emission at wavelengths generally in the range of 590-690 λ.

Examples of fluorophores that may be used are provided in U.S. Patent No. 5,866,366 to Nazarenko *et al.*, and include for instance: 4-acetamido-4'-isothiocyanatostilbene-2,2'disulfonic acid, acridine and derivatives such as acridine and acridine isothiocyanate, 5-(2'-aminoethyl)aminonaphthalene-1-sulfonic acid (EDANS), 4-amino-N-[3-vinylsulfonyl)phenyl]naphthalimide-3,5 disulfonate (Lucifer Yellow VS), N-(4-anilino-1-naphthyl)maleimide, anthranilamide, Brilliant Yellow, coumarin and derivatives such as coumarin, 7-amino-4-methylcoumarin (AMC, Coumarin 120), 7-amino-4-trifluoromethylcouluarin (Coumaran 151); cyanosine; 4',6-diaminidino-2-phenylindole (DAPI); 5', 5"-dibromopyrogallol-sulfonephthalein (Bromopyrogallol Red); 7-diethylamino-3-(4'-isothiocyanatophenyl)-4-methylcoumarin; diethylenetriamine pentaacetate; 4,4'-diisothiocyanatodihydro-stilbene-2,2'-disulfonic acid; 4,4'-diisothiocyanatostilbene-2,2'-disulfonic acid; 5-[dimethylamino]naphthalene-1-sulfonyl chloride (DNS, dansyl chloride); 4-(4'-dimethylaminophenylazo)benzoic acid (DABCYL); 4-dimethylaminophenylazophenyl-4'-isothiocyanate (DABITC); eosin and derivatives such as eosin

and eosin isothiocyanate; erythrosin and derivatives such as erythrosin B and erythrosin isothiocyanate; ethidium; fluorescein and derivatives such as 5-carboxyfluorescein (FAM), 5-(4,6-dichlorotriazin-2-yl)aminofluorescein (DTAF), 2'7'-dimethoxy-4'5'-dichloro-6-carboxyfluorescein (JOE), fluorescein, fluorescein isothiocyanate (FITC), and QFITC (XRITC); fluorescamine; IR144; IR1446; Malachite Green isothiocyanate; 4-methylumbelliferone; ortho cresolphthalein; nitrotyrosine; pararosaniline; Phenol Red; B-phycoerythrin; o-phthaldialdehyde; pyrene and derivatives such as pyrene, pyrene butyrate and succinimidyl 1-pyrene butyrate; Reactive Red 4 (Cibacron .RTM. Brilliant Red 3B-A); rhodamine and derivatives such as 6-carboxy-X-rhodamine (ROX), 6-carboxyrhodamine (R6G), lissamine rhodamine B sulfonyl chloride, rhodamine (Rhod), rhodamine B, rhodamine 123, rhodamine X isothiocyanate, sulforhodamine B, sulforhodamine 101 and sulfonyl chloride derivative of sulforhodamine 101 (Texas Red); N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA); tetramethyl rhodamine; tetramethyl rhodamine isothiocyanate (TRITC); riboflavin; rosolic acid and terbium chelate derivatives.

Other suitable fluorophores include GFP (green fluorescent protein), Lissamine™, diethylaminocoumarin, fluorescein chlorotriazinyl, naphthofluorescein, 4,7-dichlororhodamine and xanthene and derivatives thereof. Other fluorophores known to those skilled in the art may also be used.

Gene profiling array: An array containing a plurality of heterogeneous, mRNA-derived nucleic acid mixtures (also referred to as pools, targets or libraries) that have been generated from different samples (also referred to as specimens), such as different cells, tissues, or clinical samples such as biopsies. In certain embodiments, these nucleic acid mixtures proportionately reflect the abundance of each mRNA in the starting sample. Such mixtures thus contain nucleic acids that can be referred to as "expression-level reflective nucleic acid molecules" in that they reflect the amount of starting mRNA. Arrays according to the disclosure, on which are arrayed such expression-level reflective mixtures of nucleic acid molecules, are particularly useful in the detection and especially quantification of relative expression of a gene product of interest (used as a probe) in the specimens represented on the array.

The nucleic acid mixtures are spotted onto an array such that the array contains mRNA-derived mixtures (targets) from several to thousands of different cell or tissue types. These gene profiling microarrays are then probed with a single, labeled nucleic acid sequence (probe). Hybridization signals from individual spots are indicative of cell (or tissue, etc.) types that express the specific gene product that corresponds to the sequence used as a probe. This system permits the simultaneous analysis of gene product expression in a collection of specimens, and yields a "cell expression" or "tissue expression" profile for that gene product. In addition, by labeling two or more different probe sequences with different fluorescent tags, multiple genes can be profiled simultaneously on the same array.

Any procedure that results in mRNA-derived nucleic acids can be used to generate the heterogeneous mixtures of nucleic acid used. For instance, mRNA extracts could be used, as could

amplified or non-amplified cDNA preparations produced through well known techniques. It is beneficial to use an amplified nucleic acid preparation especially when only a small amount of starting material for construction of the probe is available.

By way of example only, the mixtures of target nucleic acids can be generated using the herein disclosed high fidelity mRNA-derived molecule production technique, which technique is explained more fully in the Examples (below). This method of producing target nucleic acid mixtures has certain advantages over other techniques. In particular, if the researcher is interested in information about the relative expression level of a gene in the different cell samples, it is important that the nucleic acid mixtures on the array proportionately reflect the relative abundance of the starting mRNA. The disclosed nucleic acid mixture amplification system provides this proportionate (mRNA level reflective) amplification. In addition, this system demonstrates very high fidelity amplification of mRNA nucleic acids even from very small sample amounts. Such amplification therefore can be used to produce nucleic acid mixtures for a multiple-sample, gene profiling microarray composed of nucleic acid mixtures from individual (single) source cells, fine needle aspirates, products of micro-dissection, or experimental models studying embryonic tissue or small organisms.

High throughput genomics: Application of genomic or genetic data or analysis techniques that use microarrays or other genomic technologies to rapidly identify large numbers of genes or proteins, or distinguish their structure, expression or function from normal or abnormal cells or tissues.

Human Cells: Cells obtained from a member of the species *Homo sapiens*. The cells can be obtained from any source, for example peripheral blood, urine, saliva, tissue biopsy, surgical specimen, amniocentesis samples and autopsy material. From these cells, genomic DNA, cDNA, mRNA, RNA, and/or protein can be isolated.

Hybridization: Nucleic acid molecules that are complementary to each other hybridize by hydrogen bonding, which includes Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding between complementary nucleotide units. For example, adenine and thymine are complementary nucleobases that pair through formation of hydrogen bonds. "Complementary" refers to sequence complementarity between two nucleotide units. For example, if a nucleotide unit at a certain position of an oligonucleotide is capable of hydrogen bonding with a nucleotide unit at the same position of a DNA or RNA molecule, then the oligonucleotides are complementary to each other at that position. The oligonucleotide and the DNA or RNA are complementary to each other when a sufficient number of corresponding positions in each molecule are occupied by nucleotide units which can hydrogen bond with each other.

"Specifically hybridizable" and "complementary" are terms that indicate a sufficient degree of complementarity such that stable and specific binding occurs between the oligonucleotide and the DNA or RNA target. An oligonucleotide need not be 100% complementary to its target DNA sequence to be specifically hybridizable. An oligonucleotide is specifically hybridizable when

binding of the oligonucleotide to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA, and there is a sufficient degree of complementarity to avoid non-specific binding of the oligonucleotide to non-target sequences under conditions in which specific binding is desired, for example under physiological conditions in the case of *in vivo* assays, or under conditions in which the assays are performed. Such binding is referred to as specific interference with expression of the notch protein.

Hybridization conditions resulting in particular degrees of stringency will vary depending upon the nature of the hybridization method of choice and the composition and length of the hybridizing DNA used. Generally, the temperature of hybridization and the ionic strength (especially the Na^+ concentration) of the hybridization buffer will determine the stringency of hybridization. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are discussed by Sambrook *et al.* (1989), chapters 9 and 11, herein incorporated by reference.

Isolated: An "isolated" biological component (such as a nucleic acid molecule, protein or organelle) has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs, *i.e.*, other chromosomal and extra-chromosomal DNA and RNA, proteins and organelles. Nucleic acids and proteins that have been "isolated" include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

Label: Detectable marker or reporter molecules, which can be attached to nucleic acids, for example probe molecules. Typical labels include fluorophores, radioactive isotopes, ligands, chemiluminescent agents, metal sols and colloids, and enzymes. Methods for labeling and guidance in the choice of labels useful for various purposes are discussed, *e.g.*, in Sambrook *et al.*, in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1989) and Ausubel *et al.*, in *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Intersciences (1987).

Malignant: A term describing cells that have the properties of anaplasia, invasion and metastasis.

Neoplasm: Abnormal growth of cells

Normal cells: Non-tumor, non-malignant, and non-infected cells.

Nucleic acid: A deoxyribonucleotide or ribonucleotide polymer in either single or double stranded form, and unless otherwise limited, encompassing known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides.

Nucleic acid array: An arrangement of nucleic acids (such as DNA or RNA) in assigned locations on a matrix, such as that found in cDNA arrays, or in the herein described gene profiling arrays.

Nucleic acid molecules representing genes: Any nucleic acid, for example DNA, cDNA or RNA, of any length suitable for use as a probe that is informative about the genes.

Oligonucleotide: A linear single-stranded polynucleotide sequence ranging in length from 2 to about 1,000,000 bases, for example a polynucleotide (such as DNA or RNA) which is at least 6 nucleotides, for example at least 15, 50, 100, 200, 1,000, 10,000 or even 1,000,000 nucleotides long. Oligonucleotides are often synthetic but can also be produced from naturally occurring polynucleotides.

An oligonucleotide analog refers to moieties that function similarly to oligonucleotides but have non-naturally occurring portions. For example, oligonucleotide analogs can contain non-naturally occurring portions, such as altered sugar moieties or inter-sugar linkages, such as a phosphorothioate oligodeoxynucleotide. Functional analogs of naturally occurring polynucleotides can bind to RNA or DNA, and include peptide nucleic acid (PNA) molecules. Such analog molecules may also bind to or interact with polypeptides or proteins.

Plant Cells: Cells obtained from any member of the *Plantae* Kingdom, a category which includes, for example, trees, flowering and non flowering plants, grasses, and *Arabidopsis*. The cells can be obtained from any part of the plant, for example roots, leaves, stems, or any flower part. From these cells, nucleic acid and/or protein can be isolated.

Peptide Nucleic Acid (PNA): An oligonucleotide analog with a backbone comprised of monomers coupled by amide (peptide) bonds, such as amino acid monomers joined by peptide bonds.

Probe: A molecule that can bind to or interact with one or more nucleic acid molecules. A probe, as the term is used herein, can be any nucleic acid molecule (or analog that possesses nucleic acid binding characteristics) that is used to challenge ("probe," "assay," "interrogate" or "screen") a gene profiling array, in order to determine the relative or absolute expression level of a gene in at least one spot of the array. In specific embodiments, probes may be single or double stranded nucleic acid, but will often be single-stranded DNA or RNA. In specific embodiments, the probe will be single, positive-strand nucleic acid, particularly in those embodiments wherein the mixtures of nucleic acids immobilized on the array include cDNA molecules.

Usually, a probe molecule is detectable for use in probing an array. Probes can be rendered detectable by being labeled with an independently detectable tag. The tag may be any recognizable feature that is, for example, microscopically distinguishable in shape, size, color, optical density, etc.; differently absorbing or emitting of light; chemically reactive; magnetically or electronically encoded; or in some other way detectable. Specific examples of tags are fluorescent or luminescent molecules that are attached to the probe, or radioactive monomers or molecules that can be added during or after synthesis of the probe molecule. Other tags and detection systems are known to those of skill in the art, and can be used.

Though in many embodiments a single type of probe molecule (for instance one single-stranded DNA sequence) at a time will be used to assay the array, in some embodiments, mixtures of probes will be used, for instance mixtures of two nucleic acid molecules. Such co-applied probes may be labeled with different tags, such that they can be simultaneously detected as different signals (*e.g.*, two fluorophores that emit at different wavelengths). In specific embodiments, one of these co-

applied probes will be a control probe (or probe standard), which is designed to hybridize to a known and expected sequence in one or more of the spots on the array.

Probe standard: A probe molecule for use as a control in analyzing an array. Positive probe standards include any probes that are known to interact with at least one of the nucleic acids of the array, which may be found in certain spots, or in all spots on the array, each spot containing a mixture (*e.g.*, a different mixture) of nucleic acid molecules. Negative probe standards include any probes known not to interact with any nucleic acid sequence contained in at least one mixture of nucleic acids of the array.

Such a control probe sequence could, for instance, be designed to hybridize with a so-called "housekeeping" gene, which is known to or suspected of maintaining a relatively constant expression level (or at least known to be expressed) in a plurality of cells, tissues, or conditions. Many of such "housekeeping" genes are well known; specific examples include histones, β-actin, or ribosomal subunits (either mRNA encoding for ribosomal proteins or rRNAs). Housekeeping genes can be specific for the cell type being assayed, or the species or Kingdom from which sample nucleic acid mixtures have been produced. For instance, ribulose bis-phosphate carboxylase oxygenase (RuBisCO), an enzyme involved in plant metabolism, may provide useful positive control probes for use with arrays if the nucleic acid mixtures arrayed have been derived from plant cells or tissues. Likewise, probes from the RuBisCO sequence (or any other plant-specific sequence) could provide good negative controls for gene profiling array spots that include animal-derived samples.

In some instances, as in certain of the kits that are provided herein, a probe standard will be supplied that is unlabeled. Such unlabeled probe standards can be used in a labeling reaction as a standard for comparing labeling efficiency of the test probe that is being studied. In some embodiments, labeled probe standards will be provided in the kits.

Probing: As used herein, the term "probing" refers to incubating an array with a probe molecule (usually in solution) in order to determine whether the probe molecule will hybridize to molecules immobilized on the array. Synonyms include "interrogating," "challenging," "screening" and "assaying" an array. Thus, a gene profiling array is said to be "probed" or "assayed" or "challenged" when it is incubated with a probe molecule (such as a positive, single-stranded and detectable nucleic acid molecule that corresponds to a gene of interest).

Purified: The term purified does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified nucleic acid preparation is one in which the specified protein is more enriched than the nucleic acid is in its generative environment, for instance within a cell or in a biochemical reaction chamber. A preparation of substantially pure nucleic acid may be purified such that the desired nucleic acid represents at least 50% of the total nucleic acid content of the preparation. In certain embodiments, a substantially pure nucleic acid will represent at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, or at least 95% or more of the total nucleic acid content of the preparation.

Recombinant: A recombinant nucleic acid is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination can be accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, *e.g.*, by genetic engineering techniques.

RNA: A typically linear polymer of ribonucleic acid monomers, linked by phosphodiester bonds. Naturally occurring RNA molecules fall into three classes, messenger (mRNA, which encodes proteins), ribosomal (rRNA, components of ribosomes), and transfer (tRNA, molecules responsible for transferring amino acid monomers to the ribosome during protein synthesis). Total RNA refers to a heterogeneous mixture of all three types of RNA molecules.

Sequence identity: The similarity between two nucleic acid sequences, or two amino acid sequences, is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured in terms of percentage identity (or similarity or homology); the higher the percentage, the more similar the two sequences are. Homologs or orthologs of nucleic acid or amino acid sequences will possess a relatively high degree of sequence identity when aligned using standard methods. This homology will be more significant when the orthologous proteins or nucleic acids are derived from species which are more closely related (*e.g.*, human and chimpanzee sequences), compared to species more distantly related (*e.g.*, human and *C. elegans* sequences). Typically, orthologs are at least 50% identical at the nucleotide level and at least 50% identical at the amino acid level when comparing human orthologous sequences.

Methods of alignment of sequences for comparison are well known. Various programs and alignment algorithms are described in: Smith & Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman & Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444, 1988; Higgins & Sharp, Gene, 73:237-44, 1988; Higgins & Sharp, CABIOS 5:151-3, 1989; Corpet *et al.*, *Nuc. Acids Res.* 16:10881-90, 1988; Huang *et al. Computer Appls. Biosci.* 8, 155-65, 1992; and Pearson *et al.*, *Meth. Mol. Bio.* 24:307-31, 1994. Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990, presents a detailed consideration of sequence alignment methods and homology calculations.

The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990) is available from several sources, including the National Center for Biotechnology Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. Each of these sources also provides a description of how to determine sequence identity using this program.

Homologous sequences are typically characterized by possession of at least 60%, 70%, 75%, 80%, 90%, 95% or at least 98% sequence identity counted over the full length alignment with a sequence using the NCBI Blast 2.0, gapped blastp set to default parameters. Queries searched with the blastn program are filtered with DUST (Hancock and Armstrong, *Comput. Appl. Biosci.* 10:67-70, 1994). It will be appreciated that these sequence identity ranges are provided for guidance only; it is

entirely possible that strongly significant homologs could be obtained that fall outside of the ranges provided.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences, due to the degeneracy of the genetic code. It is understood that changes in nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

One indication that two nucleic acid sequences are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

An alternative indication that two nucleic acid molecules are closely related is that the two molecules hybridize to each other under stringent conditions, as described under "specific hybridization."

Specific hybridization: Specific hybridization refers to the binding, duplexing, or hybridizing of a molecule only or substantially only to a particular nucleotide sequence when that sequence is present in a complex mixture (*e.g.* total cellular DNA or RNA). Specific hybridization may also occur under conditions of varying stringency.

Hybridization conditions resulting in particular degrees of stringency will vary depending upon the nature of the hybridization method of choice and the composition and length of the hybridizing DNA used. Generally, the temperature of hybridization and the ionic strength (especially the Na+ concentration) of the hybridization buffer will determine the stringency of hybridization. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are discussed by Sambrook *et al.* (In: *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989 ch. 9 and 11). By way of illustration only, a hybridization experiment may be performed by hybridization of a DNA molecule to a target DNA molecule which has been electrophoresed in an agarose gel and transferred to a nitrocellulose membrane by Southern blotting (Southern, *J. Mol. Biol.* 98:503, 1975), a technique well known in the art and described in Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989).

Hybridization with a target probe labeled with [^{32}P]-dCTP is generally carried out in a solution of high ionic strength such as 6 x SSC at a temperature that is 20-25° C below the melting temperature, T_m, described below. For Southern hybridization experiments where the target DNA molecule on the Southern blot contains 10 ng of DNA or more, hybridization is typically carried out for 6-8 hours using 1-2 ng/ml radiolabeled probe (of specific activity equal to 10^9 CPM/μg or greater). Following hybridization, the nitrocellulose filter is washed to remove background hybridization. The washing conditions should be as stringent as possible to remove background hybridization but to retain a specific hybridization signal.

The term T_m represents the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Because the target sequences are generally present in excess, at Tm

50% of the probes are occupied at equilibrium. The T_m of such a hybrid molecule may be estimated from the following equation (Bolton and McCarthy, *Proc. Natl. Acad. Sci. USA* 48:1390, 1962):

$$T_m = 81.5° C - 16.6(\log_{10}[Na^+]) + 0.41(\% G+C) - 0.63(\% formamide) - (600/l)$$

where l = the length of the hybrid in base pairs.

This equation is valid for concentrations of Na^+ in the range of 0.01 M to 0.4 M, and it is less accurate for calculations of Tm in solutions of higher $[Na^+]$. The equation is also primarily valid for DNAs whose G+C content is in the range of 30% to 75%, and it applies to hybrids greater than 100 nucleotides in length (the behavior of oligonucleotide probes is described in detail in Ch. 11 of Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989).

Thus, by way of example, for a 150 base pair DNA probe derived from a cDNA (with a hypothetical % GC of 45%), a calculation of hybridization conditions required to give particular stringencies may be made as follows: For this example, it is assumed that the filter will be washed in 0.3 x SSC solution following hybridization, thereby: [Na+] = 0.045 M; %GC = 45%; Formamide concentration = 0; l = 150 base pairs; $Tm=81.5 - 16.6(\log_{10}[Na+]) + (0.41 \times 45) - (600/150)$; and so Tm = 74.4° C.

The T_m of double-stranded DNA decreases by 1-1.5° C with every 1% decrease in homology (Bonner *et al.*, *J. Mol. Biol.* 81:123, 1973). Therefore, for this given example, washing the filter in 0.3 x SSC at 59.4-64.4° C will produce a stringency of hybridization equivalent to 90%; that is, DNA molecules with more than 10% sequence variation relative to the target cDNA will not hybridize. Alternatively, washing the hybridized filter in 0.3 x SSC at a temperature of 65.4-68.4° C will yield a hybridization stringency of 94%; that is, DNA molecules with more than 6% sequence variation relative to the target cDNA molecule will not hybridize. The above example is given entirely by way of theoretical illustration. It will be appreciated that other hybridization techniques may be utilized and that variations in experimental conditions will necessitate alternative calculations for stringency.

Stringent conditions may be defined as those under which DNA molecules with more than 25%, 15%, 10%, 6% or 2% sequence variation (also termed "mismatch") will not hybridize. Stringent conditions are sequence dependent and are different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5° C lower than the thermal melting point T_m for the specific sequence at a defined ionic strength and pH. An example of stringent conditions is a salt concentration of at least about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and a temperature of at least about 30° C for short probes (*e.g.* 10 to 50 nucleotides). Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. For example, conditions of 5 X SSPE (750 mM NaCl, 50 mM Na Phosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30° C are suitable for allele-specific probe hybridizations.

A perfectly matched probe has a sequence perfectly complementary to a particular target sequence. The test probe is typically perfectly complementary to a portion (subsequence) of the target sequence. The term "mismatch probe" refers to probes whose sequence is deliberately selected not to be perfectly complementary to a particular target sequence.

Transcription levels can be quantitated absolutely or relatively. Absolute quantitation can be accomplished by inclusion of known concentrations of one or more target nucleic acids (for example control nucleic acids or with a known amount the target nucleic acids themselves) and referencing the hybridization intensity of unknowns with the known target nucleic acids (for example by generation of a standard curve).

Stripping: Bound probe molecules can be stripped from an array, for instance a gene profiling array, in order to use the same array for another probe interaction analysis (*e.g.*, to determine the expression level of a different gene in the arrayed mixtures of nucleic acid molecule). Any process that will remove substantially all of the first probe molecule from the array, without also significantly removing the immobilized nucleic acid mixtures of the array, can be used. By way of example only, one method for stripping a gene profiling array is by boiling it in stripping buffer (*e.g.*, very low or no salt with 0.1% SDS), for instance for about an hour or more. The stripped array may be washed, for instance in an equilibrating or low stringency buffer, prior to incubation with another probe molecule.

Where a stripability enhancer (such as the nucleotide analog of the StripAble™ and Strip-EZ™ system from Ambion (Austin, TX)) is used, the procedures provided by the manufacturer for use with this product provide a good starting point for tailoring probing and stripping conditions for use with arrays. Addition of stripability enhancers to probes for use with arrays is optional and the disclosed arrays do not depend on them to function.

Subject: Living, multicellular vertebrate organisms, a category that includes both human and veterinary subjects for example, mammals, birds and primates.

Target: As used herein, mRNA-derived mixtures of nucleic acid molecules that are spotted onto a gene profiling array are referred to as targets. Targets on a single array can be derived from several to thousands of different cell or tissue types (more generally, from a plurality of spécimens). In certain embodiments of the arrays and methods described herein, the nucleic acid molecule mixture of the target is proportionately reflective of the mRNA levels of the starting (source) material from which the nucleic acids are derived.

In general, a target on the array should be discrete, in that signals from that target can be distinguished from signals of neighboring targets, either by the naked eye (macroarrays) or by scanning or reading by a piece of equipment or with the assistance of a microscope (microarrays).

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the

practice or testing of the present invention, suitable methods and materials are described below. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting.

II. Gene Profiling (Transcriptome) Arrays

Aspects of the present disclosure are based on the utilization of *in vitro* transcription to generate full length antisense amplified RNA (aRNA) with high fidelity, from which can then be produced amplified cDNA. Because of the high-efficiency of the amplification, minimal amounts of total (source) RNA can be amplified 80,000-fold, generating pure aRNA without losing linearity (or in other words maintaining full length mRNA-derived molecules). An outline schematic of the construction and use of a gene profiling array is shown in FIG. 4. In general, FIG. 4 shows a cell or tissue **20** that undergoes extraction **22** of a mixture of RNA **24** (*e.g.*, messenger RNA). The cell or tissue **20** may be of any phenotype, stage, histology or type (*e.g.*, different cancer cells, as well as normal cells and tissues). RNA mixture **24** (including different nucleic acid molecules, which are schematically illustrated as **24a, 24b,** and **24c**) then may be amplified at **26** to provide an amplified mixture of mRNA-derived molecules **28** (including different amplified nucleic acid molecules, schematically illustrated as **28a, 28b,** and **28c**). The amplified mixture **28** for instance can be in the form of antisense RNA (aRNA), or cDNA transcribed from the aRNA. The mixture (pool) **28** of nucleic acids (including amplified nucleic acid species **28a, 28b,** and **28c**) is then printed **30** onto the substrate **32**, for instance a microarray slide. Each spot **34** on the array then represents a unique mRNA-derived library **24** from a different specimen **20**, which will often proportionately reflect the expression levels of each of the individual mRNAs in the source.

These processes can be repeated with another specimen **40** to produce (*e.g.*, by extraction or other process **42**) a different mixture of RNA molecules **44** (including different nucleic acid molecules, schematically illustrated as **44a, 44b, 44c,** and **44d**), which can optionally be amplified **46** to produce an amplified mixture of different RNA molecules **48** (including different nucleic acid molecules, schematically illustrated as **48a, 48b, 48c,** and **48d**), reflective of the RNA mixture of the specimen from which the nucleic acid molecules were obtained **40**. As before, the amplified mixture **48** can then be applied **50** to substrate **32** to produce another spot **54** on a forming array.

Using the arrays and methods described herein, thousands of different kinds of cell types and tissues can be analyzed for gene expression simultaneously. An expression profile can be determined for each gene product of interest. This profile may include the level of expression of the gene product of interest, in terms of relative cDNA copy number and in terms of cell type or tissue distribution. In addition, multiple genes can be simultaneously profiled using probes labeled with different fluorescent labels. Since cDNA library arrays are much more stable than mRNA arrays used for Northern blots, they could be widely applied to laboratory situations without requiring stringent experimental conditions. In addition, cDNA molecules (where used) of each mixture of

nucleic acid spotted onto the array are naturally antisense and therefore bind well with sense-strand probes.

Arrays disclosed herein can be viewed as the reverse of classic cDNA microarray technology. In the disclosed gene profiling arrays, heterogeneous, mRNA-derived nucleic acid library pools **24** and **44** (referred to herein as targets or simply as nucleic acid mixtures) are generated from a plurality of samples **20** and **40**, such as different cells, tissues, or clinical samples such as biopsies. In certain embodiments, these pools proportionately reflect the abundance of each mRNA in the starting sample.

The nucleic acid mixtures, either unamplified (**24** and **44**) or amplified (**28** and **48**), are spotted **30** and **50** onto a substrate **32** to form an array, such that the array contains mixtures from several to thousands of different sources (such as cell or tissue types). It is usually better to print nucleic acid mixtures onto the same array that are in the same orientation (all mixtures positive strand, or all mixtures negative strand), so that all mixtures on the array can be probed with a singe type of probe molecule (either negative or positive strand, respectively). These gene profiling arrays can then be probed **56** (assayed) with one or more known, usually detectable (*e.g.*, labeled) nucleic acid sequence(s) **58** (referred to as a probe). Hybridization signals from individual spots (*e.g.*, signal **62** at spot **54**) on the gene profiling array are indicative of cell (or tissue) types that express the specific gene that corresponds to the sequence used as a probe. In the illustrated example shown in FIG. 4, the probe **58** represents a gene product encoded for by an RNA molecule **48d** that is present in (and was extracted from) specimen **40** but not in specimen **20**. Therefore, when the array is probed with this detectable probe molecule **58**. The probe molecule 58 is complementary to and specifically hybridizes with RNA molecule **48d** in this example. In an array that has been probed only with molecule **58**, a signal **62** is detectable only from spot **54**, which corresponds to specimen **40**.

In certain embodiments, the intensity of the hybridization signals is also measured. Hybridization intensity can be compared (between different spots on an array, between different molecule probes such as two test probes or between a test probe and a control probe or standard) in order to determine the relative expression level of the probe in individual nucleic acid mixtures. This system permits the simultaneous analysis of gene expression in the entire collection of cell/tissue samples, and yields a "cell expression" or "tissue expression" profile for that gene. In addition, by labeling two or more different probe sequences with different tags, multiple genes can be profiled simultaneously on the same array. In such examples, the two (or more) probe sequences can be used to challenge the array either simultaneously or in sequence; using different tags helps avoid stripping the array between such sequential applications.

Detection of different signal intensities is schematically depicted in FIG. 5, in which a cell or tissue **80** undergoes extraction **82** to obtain a mixture of RNA **84** (*e.g.*, messenger RNA). RNA mixture **84** (including different nucleic acid species **84a** and **84b**) may be amplified **86** to provide an amplified mixture of mRNA-derived molecules **88** (including amplified nucleic acid species **88a** and **88b**). The amplified mixture **88** (in the form of antisense RNA (aRNA), or cDNA transcribed from

the aRNA, for instance) is then printed **90** onto the substrate **92**, for instance a microarray slide. The spot **94** on the array then represents the unique mRNA-derived library **84** from a different specimen **80**, which will often proportionately reflect the expression levels of each of the individual mRNAs in the source.

These processes can be repeated with further specimens, such as specimens **100** and **110** to produce (*e.g.*, by extraction or other process **102** and **112**) a different mixture of RNA molecules **104** and **114**. Mixture **104** includes nucleic acid species **104a** and **104b**, while mixture **114** includes species **114a**, **114b**, and **114c** (in this particular example). Although each specimen is illustrated as having some unique RNA molecules, some of the RNA molecule types (*e.g.*, type "a", represented by **84a**, **104a**, and **114a**, and type "b", represented by **84b** and **114b**) are present in different mixtures. The mixtures of nucleic acids can optionally be amplified **110** and **120** (respectively) to produce an amplified mixture of different RNA-derived molecules **108** (including nucleic acid species **108a**, **108b** and **108c**), reflective of the RNA mixture of the specimen from which the nucleic acid molecules were obtained **100**, and amplified mixture **118** (including nucleic acid species **118a**, **118b**, **118c** and **118d**). As before, the amplified mixtures **108** and **118** can then be applied (**110** and **120**) to substrate **92** to produce further spots **104** and **124**, respectively.

This particular gene profiling array can then be probed **126** (assayed) with one or more known, usually detectable (*e.g.*, labeled, *) nucleic acid sequence(s) **128** (referred to as a probe). In the illustrated example shown in FIG. 5, the probe **128** represents a gene product encoded for by RNA molecule **108c** that is present in (and was extracted from) specimen **100**, and likewise represents a homologous gene product encoded for by RNA species **118c** that is present in (and was extracted from) specimen **110**. No homologous sequence was present in specimen **80**. Therefore, when the array is probed with detectable probe molecule **128**, to produce the probed array, signal **132** is detectable from spot **104** (corresponding to specimen **40**), while signal **134** of greater intensity (as indicated by the dark shading) is detectable from spot **124** (corresponding to specimen **110**). These hybridization signals of different intensity from individual spots (*e.g.*, relatively low signal **132** at spot **104** and relatively high signal **134** at spot **124**) on the gene profiling array indicate that specimens **100** and **110** express the RNA species "c," which corresponds to the sequence used as a probe, in lower and higher amounts, respectively.

Any procedure that results in mRNA-derived nucleic acids can be used to generate the heterogeneous pools (mixtures) of nucleic acid spotted onto the gene profiling arrays. For instance, mRNA extracts could be used, as could amplified or non-amplified cDNA preparations produced through known techniques.

Several characteristics of the gene profiling arrays are described below. The embodiments and examples given are meant in no way to limit the invention.

A. Choice of array members.

The target samples of interest (*e.g.*, cells and tissues) will be selected according to a wide variety of methods. For example, certain target samples of interest are well known and included in commercial culture collections, such as the ATCC (Rockville, MD). Other target samples will be identified as being of interest from journal articles, or from other investigations using high throughput technologies (*e.g.*, cDNA microarrays or Gene Chips), or with other techniques.

Any cell can serve as the source of the target nucleic acid mixtures for use in the subject arrays. For instance, an array could be assembled that reflects many cell types (or every cell type) found in an organism (such as neural, renal, gastrointestinal, cardiac, retinal, and other cell types). In other embodiments, nucleic acid mixtures derived from a certain cell type (or collection of cell types) under a variety of growth conditions (such as, different developmental stages, different nutrient conditions, different salt concentrations, and/or different temperatures) can be immobilized on one array. Alternatively, arrays can be designed that contain samples taken from cells of different species, varieties (*e.g.*, plant varieties), populations, etc. Arrays can also be produced that contain cell or tissue types from different families of cell or tissue types. Such families can be defined in various ways, including sources involved in a specific process (*e.g.*, immunological cells or tissues, or reproductive cells or tissues), sources that are in a region or organ of a subject (*e.g.*, cells or cell types found in the brain), sources known to be diseased (*e.g.*, different tumors, and more particularly samples taken from tumors at different stages of development), etc. The arrays can also be used to investigate cellular responses to drug exposure, for example by detecting differences in gene expression following *in vivo* treatment with, or *in vitro* exposure to, a drug (such as an antineoplastic agent). Arrays can also be designed to examine cellular responses to toxins in a similar fashion.

In essence, mixtures of nucleic acids from any combination or grouping of cells or tissues can be assembled together to form one or a set of gene profiling arrays for simultaneous analysis of expression of one or more genes.

Gene profiling arrays can be used to simultaneously examine gene expression in different species. Species used to produce mixed samples of nucleic acid molecules can for instance be taken from different genera, different families, different orders, different classes, different divisions, or even different kingdoms. Arrays can also be assembled that contain samples from prokaryotes (or eukaryotes), more generally.

Samples of non-human species from which specimens can be taken to prepare nucleic acid mixtures for arraying include disease organisms (*e.g.*, viruses, bacteria, parasites, etc.), research organisms (*Drosophila melanogaster*, *Caenorhabditis elegans*, *Xenopus laevis*, *Arabidopsis*, *Saccharomyces cereviseae*, *Escherichia coli*, etc.), domesticated animals (*e.g.*, cows, pigs, chickens, cats, dogs, etc.), and so forth.

Gene profiling arrays may also be used to evaluate genetic drift, population differences, progressive speciation, and other such evolution-related phenomena. Arrays can also be designed to track and study genetically-linked diseases (or other genetically determined or influenced conditions)

in families; examples of such diseases include familial predisposition to cancers (*e.g.*, breast or prostate cancers), familial hypercholestrolemia, polycystic kidney disease, Huntington disease, hereditary spherocytosis, hemophilia (and other hemoglobinopathies such as sickle cell anemia), Marfan syndrome, cystic fibrosis, Tay-Sachs disease, cystinuria, phenylketonuria, mucopolysaccharidoses, glycogen storage disease, galactosemia, homocystinuria, porphyria, Duchene muscular dystrophy. In such arrays, the mixtures of nucleic acid could be derived from cells of related family members, and could be probed with nucleic acids known or thought to be linked to the suspected genetically linked disease or condition.

Gene profiling array technology can also be used to examine progression of gene expression changes both in the same and in different tumor types, or in diseases other than neoplasia. Gene profiling arrays may be used to identify and analyze prognostic markers or markers that predict therapy outcome for various diseases or abnormal conditions, such as cancers. Arrays compiled from the nucleic acid mixtures of dozens or hundreds (or more) of tumors (for example, malignant tumors) derived from patients with known disease outcomes permit gene expression assays to be performed on those arrays, to determine important prognostic markers, or markers predicting therapy outcome, which are associated with differential or altered gene expression characteristics.

Also envisioned are arrays that are custom produced for the researcher, with an arrayed collection of nucleic acid mixtures tailored to a specific research project, research system, etc.

B. Production of array members

A purpose of the disclosed arrays and methods is to provide for analysis and detection (and optionally quantification) of gene expression in a plurality of specimens simultaneously. Thus, the array members are derived from messenger RNA (mRNA) molecules, to provide a relatively accurate indication of the level of expression of each gene in a cell. Techniques for the isolation of mRNA are well known and have been known for many years (see, for instance, Ch. 7, "*Extraction, Purification, and Analysis of Messenger RNA from Eukaryotic Cells*," Sambrook, Fritsch and Maniatis, In: *Molecular Cloning, A Laboratory Manual*, CSHL Press, 1989).

While it is possible to use extracted mRNA directly as the mixtures of nucleic acid molecules arrayed as spots on gene profiling arrays, in certain instances it is beneficial to convert the extracted mRNA into some other form, referred to generally as mRNA-derived nucleic acids, in order to for instance enhance the stability of the arrayed nucleic acids. Such mRNA-derived nucleic acids can be DNA (produced, for instance, by reverse transcription) or amplified RNA.

Likewise, in specific embodiments the extracted mRNA (or DNA derived from it) is amplified prior to the mixtures of nucleic acids being arrayed. Any amplification technique can be used, such as strand displacement amplification (as described in U.S. Patent No. 5,744,311, herein incorporated by reference), and polymerase chain reaction amplification. However, it is beneficial if the amplification method maintains the proportionality of the starting mRNA collection. Thus, preferred methods of amplified nucleic acid mixtures for use as targets will reliably produce full-

length (or predominantly full-length) nucleic acid molecules corresponding specifically to the starting mRNA species, and in approximately the same relative proportion. Such methods will produce mixtures of nucleic acid molecules that substantially reflective the expression-level of genes in the source specimen from which the sample was obtained.

This specification provides particular methods for production of mixtures of nucleic acid molecules that proportionately represent their expression, a broad description of which methods follows, and a more detailed description of which is given in Examples 1 and 2, below. One such method is also illustrated schematically in FIG. 6. The presentation of this specific embodiment is meant in no way to limit production and use of the disclosed gene profiling arrays to this method for production of pools of expression-level reflective nucleic acid molecules. Likewise, this disclosure is not meant to limit the use to which amplified mixtures of nucleic acids produced by this method are put.

As shown in FIG. 6, total RNA (which contains polyA-RNA **140**) is isolated **142** from a specimen **144** it using any standard protocol. A small amount of the total RNA, for example about 0.5 µg to about 2.0 µg, is then used as the template for first strand cDNA synthesis using reverse transcription **146**. This reaction may be primed with a generic primer **148** (for instance, an oligo-dT molecule) in order to amplify a population of mRNAs; in addition, the primer should include the antisense sequence corresponding to an RNA polymerase promoter, for instance the T7 promoter as illustrated in FIG. 6.

Second strand synthesis is initiated through template switching **150** (Matz *et al.*, *Nuc. Acids Res.* 27:1558-1560; SMART™ PCR cDNA Synthesis Kit User Manual, CLONTECH, Palo Alto, CA; WO 97/24455). When the reverse transcriptase reaches the end of the mRNA, it adds a few dC residues **152**; this is a function of terminal transferase activity of reverse transcriptase. A "template switching" oligonucleotide **154** (TS primer) containing a short string of dG residues at the 3' end is added to the mixture; it anneals to the dC string **152** on the end of the newly synthesized cDNA **156**, producing an overhang. Reverse transcriptase then switches templates to this overhang and produces a short region of duplex DNA **158**. After treatment with RNase **160** to remove the original mRNA **140**, DNA polymerase **162** is used to complete the second strand synthesis, thereby producing double strand DNA **164**. Because the only primer used to initiate second strand synthesis is the template switching primer **154**, only full-length ds cDNA **164** is produced. The RNA polymerase promoter **166** integrated in antisense with the original oligo-dT primer **148** can be used to synthesize antisense mRNA **168** (asRNA, or merely aRNA), through an RNA polymerase reaction (*e.g.*, mediated by T7 RNA polymerase **170**). Optionally, amplified cDNA mixtures **172** can be generated from this aRNA **168** through reverse transcription **174**.

A second round of amplification (not illustrated in FIG. 6) can be carried out, using a template switching primer for priming first strand synthesis and an oligo-dT primer for priming second strand synthesis. This procedure permits further amplification of the mixture of RNA derived nucleic acid molecules.. For instance, a sample of RNA extracted from a source can be amplified

once to produce a first amplified mixture of nucleic acids, and future amplified mixtures of nucleic acids produced by further amplifying using a portion of the first amplified mixture.

The production of aRNA **106** through integration of an antisense T7 promoter **86** during reverse transcription has been disclosed (see WO 99/25873; Phillips and Eberwine, *Methods* 10:283-288, 1996; and U.S. Patent No. 5,891,636 (the '636 patent). However, each of these references uses the Gulber-Hoffman (*Gene* 25:263-269, 1983) method of second strand cDNA synthesis, which employs RNase H and *E. coli* DNA polymerase I to synthesize the second strand of cDNA, rather than the template switching method employed herein. The Gulber-Hoffman system of second strand synthesis is known to tend to generate 5'-end truncated (3'-end biased) double stranded cDNA, and is therefore particularly ineffective for synthesis of cDNA from long messages (WO 97/24455).

Using template switching decreases the amount of starting RNA required, and increases full-length message production during ds cDNA production (see, Matz *et al., Nuc. Acids Res.* 27:1558-1560, 1999; WO 97/24455). The "SMART™" system (Switch Mechanism At the 5' end of RNA Templates), offered by CLONTECH Laboratories (Palo Alto, CA), is a commercially available template switching system recommended for use in library construction.

The described method of producing mixtures of mRNA-reflective nucleic acid molecules requires substantially less starting material (0.5 µg of total RNA) than required by the method of Lockhart *et al.* (*Nat. Biotech.* 14:1675-1680, 1996), which requires about 1 µg of polyA-RNA, or about 200 times more material than is necessary for the disclosed system. Template switching also amplifies only full-length cDNAs, in contrast to the Gulber-Hoffman synthesis, which can produce shortened cDNA (through the effect known as 3' bias). In addition, template switching is carried out at a higher temperature (75° C) than the Gulber-Hoffman synthesis (37° C), which reduces non-specific priming and thereby increases the fidelity of the amplification process disclosed herein.

Mixtures of amplified nucleic acid molecules that reflect the mRNA level of the specimen from which the source RNA was obtained, produced as described above, can be used for other purposes than as targets in the herein disclosed gene profiling arrays. For instance, such mixtures of nucleic acid molecules can be labeled and used as a "target" in for analysis of a conventional cDNA microarray. Because the disclosed method requires very little starting material, this application would open up conventional cDNA microarray analysis to entire new fields of research, especially those in which the source material was heretofore too scarce to permit cDNA array analysis (*e.g.*, for samples acquired by fine needle aspirates or micro-dissection, or experimental models studying embryonic tissue or small organisms). Also encompassed are these other uses of the herein disclosed nucleic acid amplification technique.

C. Choice of array format and structure

Gene profiling arrays may vary significantly in their structure, composition, and intended functionality. The disclosed array system is amenable to use in either a macroarray or a microarray format, or a combination thereof. Such arrays can include, for example, at least 50, 100, 150, 200,

500, 1000, or 5000 or more array elements (such as spots). In the case of macro-format gene profiling arrays, no additional sophisticated equipment is usually required to detect the bound (hybridized) probe on the gene profiling array, though quantification may be assisted by known automated scanning and/or quantification techniques and equipment.

Examples of substrates for the disclosed arrays include glass (*e.g.*, functionalized glass), Si, Ge, GaAs, GaP, SiO_2, SiN_4, modified silicon nitrocellulose, polyvinylidene fluoride, polystyrene, polytetrafluoroethylene, polycarbonate, nylon, fiber, or combinations thereof. Array substrates can be stiff and relatively inflexible (*e.g.*, glass or a supported membrane) or flexible (such as a polymer membrane). One commercially available microarray system that can be used with the arrays is the FAST™ slides system (Schleicher & Schuell, Dassel, Germany), which incorporates a patch of polymer on the surface of a glass slide.

Macro-format gene profiling arrays are often arrayed on polymer membranes, either supported or not, and can be of any size, but typically will be greater than a square centimeter. Other examples of macroarray substrates include glass, fiber, plastic and metal. Macroarrays are generally used when the number of mixtures of nucleic acids (pools) in the target set is relatively small, on the order of tens to hundreds of samples, however macroarrays with a larger number of array elements can be used on large substrates. Spot arrangement on the macroarray is such that individual spots can be distinguished from each other when the sample is read; typically, the diameter of the spot is about equal to the spacing between individual dots.

Sample spots on macroarrays are of a size large enough to permit their detection without the assistance of a microscope or other sophisticated enlargement equipment. Thus, spots may be as small as about 0.1 mm across, with a separation of about the same distance, and can be larger. Larger sample spots on macroarrays, for example, may be about 0.5, 1, 2, 3, 5, 7, or 10 mm across. Even larger spots may be larger than 10 mm (1 cm) across, in certain specific embodiments. The array size will in general be correlated the size of the sample spots applied to the array, in that larger spots will usually be found on larger arrays, while smaller spots may be found on smaller arrays. This correlation is not necessary, though.

In microarray-format gene profiling arrays, a common feature is the small size of the target array, for example an area of about a squared centimeter (1 cm^2) or less. A squared centimeter (for example, a square of dimensions 1 cm by 1 cm) is large enough to contain over 2,500 individual target spots, if each spot has a diameter of 0.1 mm and spots are separated by 0.1 mm from each other. A two-fold reduction in spot diameter and separation can allow for 10,000 such spots in the same array, and an additional halving of these dimensions would allow for 40,000 spots. Using microfabrication technologies, such as photolithography, pioneered by the computer industry, spot sizes of less than 0.01 mm are feasible, potentially providing for over a quarter of a million different target sites. The power of microarray-format gene profiling arrays resides not only in the number of different mixtures of nucleic acid that can be probed simultaneously, but also in how little starting

material is need for the target. Spots on a microarray will generally be no larger than about 1 mm by 1 mm.

The amount of target nucleic acid mixture that is applied to each address of an array will be largely dependent on the array format used. For instance, microarrays will generally have less nucleic acid applied at each address than will macroarrays. By way of example, individual targets on a macroarray can be applied in the amount of about 0.5 µg or greater, for instance about 1 µg, about 3 µg, about 5 µg, about 7.5 µg, about 10 µg, about 15 µg or more. In contrast, samples applied to individual spots on a gene profiling microarray will usually be less than 1 µg in each spot, for instance, about 0.5 µg, about 0.1 µg, about 0.08 µg, about 0.05 µg, about 0.01 µg or less. In certain applications, each spot on the array may contain as little as 0.005 µg of nucleic acid mixture. Where all of the nucleic acids in each mixture are single stranded (*e.g.*, where the nucleic acid mixture is a mixture of amplified, single-stranded cDNA molecules), no material will be lost in having to denature the array before it can be probed.

In addition, the surface area of sample application for each "spot" will influence what amount of nucleic acid mixture is immobilized on the array surface. Thus, a larger spot (having a greater surface area) will generally accept or require a greater amount of target molecule than a smaller sample spot (having a smaller surface area).

Characteristics of the target nucleic acids in the mixtures (*e.g.*, the length of the cDNA molecule, its primary and secondary structure, its binding characteristics in relation to the array substrate, etc.) will influence how much of each target mixture is applied to an array. Optimal amounts of target mixtures for application to an array can be easily determined, for instance by applying varying amounts of the target mixture(s) to an array surface and probing the array with a probe known to interact with at least one nucleic acid molecule within that target mixture. In this manner, it is possible to empirically determine a range of target nucleic acid mixture amounts that will produce interpretable results with any collection of desired nucleic acid mixtures.

Another way to describe an array is its density, for example the number of samples in a certain specified surface area. For macroarrays, array density will usually be between about one target location per squared decimeter (dm^2) (for example, one target address in a 10 cm by 10 cm region of the array substrate) to about 50 targets per cm^2 (for example, 50 targets within a 1 cm by 1 cm region of the substrate). For microarrays, array density will usually be one target location per cm^2 or more, for instance about 50, about 100, about 200, about 300, about 400, about 500, about 1000, about 1500, about 2,500, about 5,000, about 10,000, about 50,000, about 100,000 or more targets per cm^2.

D. Application of targets to arrays

After production and appropriate purification (as discussed above), nucleic acid target mixtures can be deposited onto the array using any of a variety of techniques. Though the nucleic acids being deposited are different than in traditional microarray technology, the techniques described

for these traditional systems are equally applicable to deposition of the herein disclosed nucleic acid preparation to gene profiling arrays. For instance, arrays can be formed on non-porous surfaces (such as glass) by robotic micropipetting of nanoliter quantities of DNA to predetermined positions on a non-porous glass surface (as in Schena *et al.*, *Science* 270:467-470, 1995, and WO 95/35505). This is a "spotting" technique. Generally, in a spotting technique, the target molecules are delivered by directly depositing (rather than flowing) relatively small quantities of them in selected regions. For instance, a dispenser can move from address to address, depositing only as much target as necessary at each stop. Typical dispensers include an ink-jet printer or a micropipette to deliver the target in solution to the substrate, and a robotic system to control the position of the micropipette with respect to the substrate. In other embodiments, the dispenser includes a series of tubes, a manifold, an array of pipettes, or the like so that the target polypeptides can be delivered to the reaction regions simultaneously.

Usually, the target nucleic acid mixtures are deposited on the array substrate in such a way that they are substantially irreversibly bound to the array. For example, a target may be bound such that no more than 30% of the molecules in the mixtures on the array at the end of the binding process can be washed off using buffers of the gene profiling array system (*e.g.*, low or high stringency wash buffers or stripping buffers). In other embodiments, no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, no more than 3%, or no more than 1% of the nucleic acids on the array at the end of the binding process can be washed off using buffers of the gene profiling array system.

Depending on the array substrate used, the substrate alone may substantially irreversibly bind the target nucleic acids without further linking being necessary (*e.g.*, nitrocellulose and PVDF membranes). In other instances, a linking or binding process must be performed to ensure binding of the polypeptides. Examples of linking processes are known to those of skill in the art, as are the substrates that require such a linking process in order to bind polypeptide molecules. For instance, deposited nucleic acid molecules may be coupled to the solid support by electrostatic interactions with a coating film of a polycationic polymer such as poly-L-lysine (WO 95/35505), or covalently bound to the solid support. The target nucleic acids optionally may be attached to the array substrate through linker molecules.

In certain embodiments, the non-sample regions of the array surface (those regions of the array surface that do not contain target molecules) are blocked in order to prevent or inhibit non-specific binding of probe molecules directly to the array surface.

E. Choice of probe molecule(s)

Many different probe molecules can be used with the arrays and methods disclosed herein. Probes can be selected, for example, based on the needs of an individual investigator. Since the spots on the array will contain nucleic acid corresponding to substantially every expressed sequence within

the specimens chosen, probes for use with the gene profiling arrays can represent any gene product of interest.

A hybridization probe for use in an array produced according to this disclosure may be referred to as a sequence "representing" a particular gene or gene product. A sequence "representing" a particular gene product is one that will specifically hybridize to a nucleic acid molecule encoding that gene product, thereby permitting identification of that gene product. A sequence representing a particular gene product may include an entire cDNA sequence (or the corresponding genomic gene sequence) or less than an entire cDNA sequence. For example, the probe may include an oligonucleotide comprising a minimum specified number of consecutive bases of a selected gene that is differentially expressed. Oligonucleotides as short as 8-10 consecutive bases of a cDNA will be effective to produce meaningful gene expression data using microarray technology. For example, a nine base oligonucleotide can distinguish 262,144 transcripts (4^9). However, for enhanced specificity of hybridization, longer oligonucleotides may be employed, such as at least 10, 15, 20, 25, 30, 50, 50 or more consecutive bases of a cDNA. Other examples of probe molecules that are shorter than the full length of the subject cDNA include individual exons of the gene sequence of interest, ESTs from within the gene sequence, or regions of the nucleotide sequence of interest that encode conserved regions within the encoded proteins (and thereby may be useful to examine the expression of related proteins). In the latter example, it will be advantageous in certain embodiments to produce a collection of degenerate probe molecules; production of such degenerate probes is known.

Furthermore, a probe "representing" a particular gene product need not be a complete match. While probes that share 100% sequence identity over their entire length to the corresponding cDNA sequence will typically provide enhanced specificity of hybridization, probes that share less than 100% sequence identity may also be useful in such microarray applications. Typically, such probes will share at least 70% sequence identity with the corresponding cDNA, but probes sharing at least 75%, 80%, 85%, 90%, 95%, 97%, 98%, and 99% sequence identity may be utilized to achieve enhanced specificity. Probes can also be selected based on their specific complementarity or degree of hybridization to the target sequence.

In many embodiments, it is beneficial also to prepare a probe molecule for use as a control in analyzing the gene profiling array. Positive probe standards include any probes that are known to interact with at least one of the nucleic acids of the array, which may be found in certain spots, or in all spots on the array. Negative probe standards include any probes known not to interact with any nucleic acid sequence contained in at least one mixture of nucleic acids (contained in a spot) of the array. Control probe sequences could, for instance, be designed to hybridize with a so-called "housekeeping" gene, which is known to or suspected of maintaining a relatively constant expression level (or at least known to be expressed) in a plurality of cells, tissues, or conditions. Many of such "housekeeping" genes are well known; specific examples include histones, β-actin, or ribosomal subunits (either mRNA encoding for ribosomal proteins or rRNAs).

F. *Labeling and detection of probe molecule(s)*

Usually, probe molecules used to assay the disclosed gene profiling arrays are detectable. Probes can be rendered detectable by being labeled with an independently detectable tag or other reporter molecule. Such tags include fluorescent or luminescent molecules that are attached to the probe, or radioactive monomers or other detectable molecules that can be added during or after synthesis of the probe molecule.

Labeling different probes with different tags, each of which can be detected simultaneously (*e.g.*, two fluorophores that fluoresce at different wavelengths) enables simultaneous detection of hybridization of two or more probes on the nucleic acid mixtures of an array. Multiple-label challenges to an array can also be used to provide an internal control. For competitive binding assays, however, only one of the probes needs to be detectable. The detectable label (e.g., the fluorophore) may be incorporated during synthesis of the probe.

It will be appreciated that the color of the labels used is not critical, so long as the emission wavelength of the different fluorophores used can be resolved, and can be used to measure differential expression. Other fluorophores or labels can be used to practice the disclosed methods.

Typical experiments involve either single-color fluorescence hybridization to measure the levels of expression of a single gene in all of the arrayed specimens, or two-color fluorescence hybridization to examine the relative expression of genes of two different genes simultaneously, or to provide an internal (*e.g.*, quantitative) control for the detection of expression of a single gene.

For single-color fluorescence hybridization experiments, a probe molecule corresponding to a gene of interest is produced. The probe is labeled, for example using a fluorescent dye such as Cy3 or Cy5 (Amersham Pharmacia Biotech, Piscataway, NJ), or any other fluorophore or label. The label can be incorporated directly during synthesis. The probe is then hybridized to the array. Following washing to remove non-specifically bound probe, the array is scanned for fluorescent emission following laser excitation, and the intensity of each fluorescing spot is measured. The intensity of each spot is approximately proportional to the expression of the gene (corresponding to the probe) in each nucleic acid mixture contained within a spot on the array. This data provides an indication of the expression of a particular gene (corresponding to the labeled probe) in the specimens (*e.g.*, cells or tissues) from which the mixtures of nucleic acids were prepared.

For two-color fluorescence hybridization experiments, two probe molecules are produced and labeled as described above, except that each probe is labeled with a different fluorescent label, each of which fluoresces at a different wavelength (for example, one sample may be labeled with Cy3 and the other with Cy5). After the two probe preparations are labeled, they are mixed together and hybridized to a single array. Alternatively, they can be applied to the single array sequentially in certain embodiments. After washing, the array is scanned using two fluorescence channels. Because the two fluorescent labels are selected such that their emission spectra do not overlap, the signal of each of the two fluors can be measured for each of the probes. The absolute levels of intensity for

each probe in an array is approximately proportional to the expression of the gene in the sample examined, and the ratio of the two fluor intensities indicates the relative expression of a gene in the two different samples.

Where one of the probes used in a two-color experiment is used as a control, and is directed toward a "housekeeping" gene, its signal intensity at each spot can be used to normalize the hybridization signal intensity of the test probe at each corresponding spot.

G. Optional additional elements of probe molecules

In addition to adding label during the synthesis, it is also optionally possible to add other elements that enhance or alter the activity of the probe. By way of example, one such possible addition is an altered nucleic acid residue that renders the probe molecule easy to degrade under certain circumstances. On such altered nucleic acid residue can be purchased from Ambion (Austin, TX) under the name of the StripAble™ and Strip-EZ™ system. This system enhances the stripability of a probed array by providing for the degradation of probe molecule under relatively gentle conditions (detailed in the Strip-EZ™ protocol) that substantially reduce the loss of immobilized target nucleotide during stripping procedures. Incorporation of this nucleic acid analog, or other similarly functional analogs, into probes can increase the life span of the array and enhance the detectability of gene expression signals using probes to several more gene products. Such additional elements are optional and the invention does not depend on them to function.

H. Computer assisted (automated) detection and analysis of array

The data generated by assaying a gene profiling array can be analyzed using known computerized systems. For instance, the array can be read by a computerized "reader" or scanner and quantification of the binding of probe to individual addresses on the array carried out using computer algorithms. Likewise, where a control probe has been used, computer algorithms can be used to normalize the hybridization signals in the different spots of the array. Such analyses of an array can be referred to as "automated detection" in that the data is being gathered by an automated reader system.

In the case of labels that emit detectable electromagnetic wave or particles, the emitted light (*e.g.*, fluorescence or luminescence) or radioactivity can be detected by very sensitive cameras, confocal scanners, image analysis devices, radioactive film or a Phosphoimager, which capture the signals (such as a color image) from the array. A computer with image analysis software detects this image, and analyzes the intensity of the signal for each probe location in the array. Signals can be compared between spots on a single array, or between arrays (such as a single array that is sequentially probed with multiple different probe molecules), or between the labels of different probes on a single array.

Computer algorithms can also be used for comparison between spots on a single array or on multiple arrays. In addition, the data from an array can be stored in a computer readable form.

Certain examples of automated array readers (scanners) will be controlled by a computer and software programmed to direct the individual components of the reader (*e.g.*, mechanical components such as motors, analysis components such as signal interpretation and background subtraction). Optionally software may also be provided to control a graphic user interface and one or more systems for sorting, categorizing, storing, analyzing, or otherwise processing the data output of the reader.

To "read" an array, an array that has been assayed with a detectable probe to produce binding (*e.g.*, a binding pattern) can be placed into (or onto, or below, etc., depending on the location of the detector system) the reader and a detectable signal indicative of probe binding detected by the reader. Those addresses at which the probe has bound to an immobilized nucleic acid mixture provide a detectable signal, *e.g.*, in the form of electromagnetic radiation. These detectable signals could be associated with an address identifier signal, identifying the site of the "positive" hybridized spot. The reader gathers information from each of the addresses, associates it with the address identifier signal, and recognizes addresses with a detectable signal as distinct from those not producing such a signal. Certain readers are also capable of detecting intermediate levels of signal, between no signal at all and a high signal, such that quantification of signals at individual addresses is enabled.

Certain readers that can be used to collect data from the arrays, especially those that have been probed using a fluorescently tagged molecule, will include a light source for optical radiation emission. The wavelength of the excitation light will usually be in the UV or visible range, but in some situations may be extended into the infra-red range. A beam splitter can direct the reader-emitted excitation beam into the object lens, which for instance may be mounted such that it can move in the x, y and z directions in relation to the surface of the array substrate. The objective lens focuses the excitation light onto the array, and more particularly onto the (polypeptide) targets on the array. Light at longer wavelengths than the excitation light is emitted from addresses on the array that contain fluorescently-labeled probe molecules (*i.e.*, those addresses containing a nucleic acid molecule within a spot containing a nucleic acid molecule to which the probe binds).

In certain embodiments, the array may be movably disposed within the reader as it is being read, such that the array itself moves (for instance, rotates) while the reader detects information from each address. Alternatively, the array may be stationary within the reader while the reader detection system moves across or above or around the array to detect information from the addresses of the array. Specific movable-format array readers are known and described, for instance in U.S. Patent No. 5, 922,617, hereby incorporated in its entirety by reference. Examples of methods for generating optical data storage focusing and tracking signals are also known (see, for example, U.S. Pat. No. 5,461,599, hereby incorporated in its entirety by reference).

For the electronics and computer control, a detector (*e.g.*, a photomultiplier tube, avalanche detector, Si diode, or other detector having a high quantum efficiency and low noise) converts the optical radiation into an electronic signal. An op-amp first amplifies the detected signal and then an

analog-to-digital converter digitizes the signal into binary numbers, which are then collected by a computer.

I. Gene Profiling Array Kits

Gene profiling arrays as disclosed herein can be supplied in the form of a kit for use in gene expression analyses. In such a kit, at least one gene profiling array is provided. The kit also includes instructions, usually written instructions, to assist the user in probing the array. Such instructions can optionally be provided on a computer readable medium.

Kits may additionally include one or more buffers for use during assay of the provided array. For instance, such buffers may include a low stringency wash, a high stringency wash, and/or a stripping solution. These buffers may be provided in bulk, where each container of buffer is large enough to hold sufficient buffer for several probing or washing or stripping procedures. Alternatively, the buffers can be provided in pre-measured aliquots, which would be tailored to the size and style of array included in the kit.

Certain kits may also provide one or more containers in which to carry out array-probing reactions.

Kits may in addition include either labeled or unlabeled control probe molecules, to provide for internal tests of either the labeling procedure or probing of the gene profiling array, or both. The control probe molecules may be provided suspended in an aqueous solution or as a freeze-dried or lyophilized powder, for instance. The container(s) in which the controls are supplied can be any conventional container that is capable of holding the supplied form, for instance, microfuge tubes, ampoules, or bottles. In some applications, control probes may be provided in pre-measured single use amounts in individual, typically disposable, tubes or equivalent containers.

The amount of each control probe supplied in the kit can be any particular amount, depending for instance on the market to which the product is directed. For instance, if the kit is adapted for research or clinical use, sufficient control probe(s) likely will be provided to perform several controlled analyses of the array. Likewise, where multiple control probes are provided in one kit, the specific probes provided will be tailored to the market and the accompanying kit. In certain embodiments, a plurality of different control probes will be provided in a single kit, each control probe being from a different type of specimen found on an associated array (*e.g.*, in a kit that provides both eukaryotic and prokaryotic specimens, a prokaryote-specific control probe and a separate eukaryote-specific control probe may be provided).

In some embodiments, kits may also include the reagents necessary to carry out one or more probe-labeling reactions. The specific reagents included will be chosen in order to satisfy the end user's needs, depending on the type of probe molecule (*e.g.*, DNA or RNA) and the method of labeling (*e.g.*, radiolabel incorporated during probe synthesis, attachable fluorescent tag, etc.).

Further kits are provided for the labeling of probe molecules for use in assaying arrays provided herein. Such kits may optionally include an array to be assayed by the so labeled probe molecules. Other components of the kit are largely as described above for kits for the assaying of gene profiling arrays.

III. Examples

The following examples are provided to illustrate certain particular features and/or embodiments. These examples should not be construed to limit the invention to the particular features or embodiments described.

Example 1: Preparation of High Fidelity Array Targets

A375 melanoma and ML-1 lymphoid cell lines were obtained from the American Type Culture Collection (Rockville, MD) and the National Human Genome Research Institute respectively and maintained in RPMI supplemented with 10% fetal calf serum (Biofuilds, Rockville, MD). Total RNA was isolated using RNeasy midi kits (QIAGEN, Valencia, CA) and refined using TRIZOL reagent (Gibco-BRL, Gaithersburg, MD). The mRNA was purified from total RNA using Oligotex mRNA isolation kit (QIAGEN). RNA concentrations were determined by OD-260 reading in 50 mM sodium hydroxide (GeneQuant, Clamart Cedex, France).

The aRNA was prepared from total RNA in 9 µl DEP-C treated H_2O containing 1 µl (1 µg/µl) oligo-dT -T7 primer (5' AAA CGA CGG CCA GTG AAT TGT AAT ACG ACT CAC TAT AGG CGC TTT TTT TTT TTT T 3', SEQ ID NO: 1). Total RNA was denatured at 70° C for 3 minutes and primed while cooling to room temperature. T7 bacteria phage promoter was incorporated into cDNA synthesis in a reverse transcription (RT) reaction by adding 4 µl of first strand-reaction buffer, 2 µl 0.1M DTT (Gibco-BRL), 2 µl 10 mM dNTP, 1 µl RNAsin (Promega, Madison, WI), 1 µl (1 µg/µl) template switch primer (5'-AAG CAG TGG TAT CAA CGC AGA GTA CGC GGG-3', SEQ ID NO: 2) (CLONTECH, Palo Alto, CA) and 2 µl Superscript-II reverse transcriptase (Gibco-BRL). cDNA synthesis was carried out at 42° C for at least 1 hour. Full-length ds cDNA was synthesized by adding 106 µl of DNAse-free water, 15 µl Advantage PCR buffer (CLONTECH), 3 µl 10 mM dNTP, 1 µl RNase-H (Promega), 3 µl Advantage cDNA Polymerase (CLONTECH). The following temperature cycle was used: two minutes at 37° C for RNA digestion, 3 minutes at 94° C for denaturation, 3 minutes at 65° C for priming and 30 minutes at 75° C for extension. Reactions were terminated by incubation in 7.5 µl 1M NaOH with 2 mM EDTA at 65° C for 10 minutes. cDNA was phenol-chloroform-isoamyl extracted and ethanol precipitated in the presence of 0.1 µg linear acrylamide (0.1 µg/µl, Ambion, Austin, TX). cDNA re-suspended in 60 µl DEPC H_2O was passed through a Bio-6 chromatography column (Bio-Rad, Cambridge, MA) that had previously been washed three times with 700 µl DEPC treated H_2O. Samples were lyophilized to 16 µl. For the first round of amplification, 16 µl of purified full-length ds-cDNA was incubated with 4 µl of each 75 mM NTP (ATP, GTP, CTP and UTP), 4 µl of 10X reaction buffer and 4 µl of transcription enzyme mixture (T7 Megascript Kit 1334, Ambion) in 40 µl volume at 37° C for 5 to 6

hours. RNA recovery and removal of template DNA was achieved by TRIZOL purification. 1.3 µg of aRNA prepared from 31 ng and 0.65 µg of aRNA prepared from 10 ng source total RNA were reverse transcribed into cDNA using 2 µg of random hexamer with 5 µl first strand buffer, 2 µl 0.1M DTT, 1 µl RNAsin, 2 µl of 10 mM dNTP and 2 µl of Superscript II (SII). The reaction mixture was heated to 65° C for 10 minutes before adding SII then synthesis was continued at 42° C for 1 hour. Second strand cDNA synthesis was initiated by 1 µg oligo dT-T7 primer in the conditions used in the first round. *In vitro* transcription of aRNA was carried out as for the first round.

Fifty µg (for Cy3 labeling) or 100 µg (for Cy5 labeling) total RNA and 3 µg aRNA or non-amplified mRNA were labeled in a reverse transcription reaction by using 8 µg of random hexamer primer in the presence of Cy3 or Cy5 labeled dUTP (Amersham, Piscataway, NJ) using Superscript II (Gibco-BRL). Reaction products were purified in Bio-6 chromatography column followed by Microcon concentration. (Purified and labeled cDNA assaying molecule in 20 µl containing 2.6 µl 20 x SSC, 8 µg of poly (dA), 4 µg yeast tRNA and 10 µg of human Cot I DNA (Gibco, BRL, Life Technologies, Rockville, MD).) Prior to hybridization, the mixture was heated to 99° C for 2 minutes and then cooled to room temperature. At that point 0.46 µl of 10% SDS were added. Hybridization was carried out at 65° C for 12 to 18 hours in water bath. Prior to scanning, slides were washed in 2 x SSC with 0.1% SDS for 2 minutes, 1 x SSC, 0.2 x SSC and 0.05 x SSC sequentially for 1 minute each.

Then 2008 named cDNAs were spotted onto poly-L-lysine-coated slides using an OmniGrid arrayer (GeneMachines; San Carlos, CA). Hybridized arrays were scanned at 10-µm resolution on a GenePix 4000 scanner (Axon Instruments, Inc.; Foster City, CA) at variable PMT voltage to obtain maximal signal intensities with < 1% microarray probe saturation. Resulting TIFF images were analyzed via ArraySuite software (National Human Genome Research Institute; Bethesda, MD).

Example 2: Analysis of High Fidelity Array Nucleic Acid Target

One round of amplification yielded ~10^3-and two rounds ~10^5-fold the estimated amount of starting mRNA. Random bias resulting from RNA amplification or non-specific hybridization was assessed by hybridizing differentially labeled aRNA-based microarray targets from the same melanoma line to 2008 gene human microarrays (NCI-OncoChip). Scatter plots of Cy3 (green) versus Cy5 (red) signal reproducibly revealed a strong linear relationship ($R^2 = 0.99$) (FIG. 1, top panel). Similar linearity was observed with aRNA from a renal cancer line ($R^2 = 0.96$). To assess systematic bias introduced by aRNA amplification, the expression profile of labeled aRNA-based microarray targets was compared to that of conventional total and poly(A) RNA-based microarray targets by identifying differentially expressed genes from two different sources (A375 and ML-1) (FIG. 1, bottom panel). Truly differentially expressed genes were considered those resulting in highly reproducible "outliers" in four consecutive total RNA-based arrays at optimized microarray target concentration (100 µg for Cy3 and 50 µg for Cy5 microarray targets). Outliers were defined as genes whose array spots exhibit Cy3/Cy5 ratios significantly different from 1.0 at a 99.0% confidence

level (cutoff ratio ranged from 1.7 to 2.1). To exclude labeling biases, total RNA-based microarray targets from either cell line were labeled with the reciprocal fluorochrome in every other duplicate experiment. Therefore, a green spot on one array would be red in the reciprocal. True (concordant) outliers were those that were positives using reciprocal fluorochrome and reproducible using the same fluorochrome. Results were analyzed using the hierarchical clustering technique of Eisen *et al.* (*Proc. Natl. Acad. Sci. USA* 95:14863-14868, 1998). Outliers were ranked into mutually exclusive confidence groups. The "4/4 match" group consisted of concordant spots (N = 267) in all 4 hybridizations. The "3/4 match" group represented concordance in 3 hybridizations (N = 69). The "2/4 match rec" group contained only reciprocal outliers (N = 12) and the "2/4 match rep" group reproducible outliers (N = 311) appearing twice in the four consecutive arrays but not in the reciprocal fluorochrome experiments. The fourth group (2/4 match rep) was believed to represent genes whose measurement of expression was confounded by labeling bias affecting low transcript levels in which background fluorescence intensity was higher with one but not the other dye.

Outliers identified by aRNA-based hybridization were matched to the four confidence groups (FIG. 2A).. Eighty-five to ninety-two per cent of outliers identified by the aRNA amplified from 0.25 to 3.0 μg source RNA reproducibly matched "true outliers" identified by total RNA. The level of concordance was identical comparing an additional hybridization using total RNA or poly(A)-RNA. Detection of true outliers degenerated in aRNA amplified from 0.125 to 0.031 μg total RNA (30 to 70%). However, a second amplification restored concordance in aRNA from 0.031 to 0.010 μg total RNA (80 to 85%). To visually demonstrate the level of outlier concordance a high-stringency filter was applied (Cy3/Cy5 or Cy5/Cy3 ratios above 3, fluorescence intensity >300 in one channel unless the other channel was >1,000 and a spot size of $\leq$50 pixels). Genes satisfying these requirements in at least five experiments were clustered (Eisen *et al.*, *Proc. Natl. Acad. Sci. USA* 95:14863-14868, 1998) (FIG. 2B). Clustering revealed 251 outliers with strong concordance that decreased with reducing amounts of source RNA and could be re-established by a second amplification of low source material.

To identify false positives, a more tolerant filter (Cy5/Cy3 or Cy3/Cy5 above 3, fluorescence intensity >150 in one channel in any of the experiments) was applied allowing visualization of less reproducible outliers. Approximately 250 false positives biasing the Cy5 channel (blue bar in FIG. 3A) emerged with aRNA from <0.125 μg of source total RNA. These false outliers were not detected with total RNA or aRNA from 3.0 μg to 0.25 μg.

Not meaning to be bound by a single theory, it was postulated that this Cy5 bias was related to differential optical detection of the red and green fluorochrome at low microarray target concentration. One round of amplification from 62 to 31 ng total RNA yielded quantities of aRNA (1.3 to 2.0 μg) below the standard concentration (3 μg) of molecules used for assaying aRNA or poly(A)-RNA-based arrays. Consequently, lower amounts of labeled assaying molecule were used in these arrays decreasing fluorescence particularly in low abundance transcripts. Re-amplification of aRNA from 0.031 to 0.010 μg total RNA permitted aRNA-based hybridization with optimal

microarray target concentration (3 μg) and restored the ability to detect outliers in each confidence group with percentages comparable to the 0.25 to 3.0 μg aRNA set. Furthermore, false positive signals in the Cy5 channel were suppressed (FIG. 3A) suggesting that the Cy5 bias was not due to molecular anomalies from RNA amplification but to a remediable post-amplification artifact.

The number of experimental outliers discordant from the four confidence groups was summarized as percentage of the total number of genes on the array (FIG. 3B). This parameter is a reliable measure of non-reproducibility and was 4.5% when using labeled total RNA-based microarray targets. The percentage of non-reproducible outliers noted with aRNA-based hybridizations from 0.25 μg - 3.0 μg source RNA ranged from 3 to 6% similarly to total RNA-based arrays. This measure of non-reproducibility increased in arrays using aRNA from 0.031 to 0.125 μg source total RNA but was reduced to baseline levels by a second round of aRNA amplification.

Analysis

In vitro transcription has been utilized for differential gene expression studies (Lockhart *et al., Nat, Biotechnol.* 14:1675-1680, 1996; Luo *et al., Nat. Med.* 5:117-122, 1999; Van Gelder *et al., Proc. Natl. Acad. Sci. USA* 87:1663-1667, 1990; Eberwine *et al., Proc. Natl. Acad. Sci. USA* 89:3010-3014, 1992; Kacharmina *et al., Methods Enzymol* 303:3-19, 1999). However, these studies have estimated the linearity and reproducibility of poly(A)-RNA amplification in a limited number of genes by Northern Blot or *in situ* hybridization (Lockhart *et al., Nat, Biotechnol.* 14:1675-1680, 1996; Luo *et al., Nat Med* 5:117-122, 1999; Van Gelder *et al., Proc. Natl. Acad. Sci. USA* 87:1663-1667, 1990; Eberwine *et al., Proc. Natl. Acad. Sci. USA* 89:3010-3014, 1992; Kacharmina *et al., Methods Enzymol* 303:3-19, 1999). Conventional anti-sense mRNA amplification can introduce biases in the amplified product because of a possible 5' under-representation and because of the low stringency temperature applied during double stranded cDNA (ds-cDNA) synthesis. In the above-described procedure, a modification of conventional anti-sense mRNA amplification (Kacharmina *et al., Methods Enzymol* 303:3-19, 1999) exploiting template-switching effect at 5' end (Matz *et al., Nucleic Acids Res* 27:1558-1560, 1999) ensured the generation of full-length ds-cDNA. Furthermore, the template-switching primer-dependent second strand cDNA synthesis occurs at 75° C. Thus, this modification overcomes potential 3' bias (useful when unmapped sequences are used for array spotting) and enhances sequence specificity by high-temperature cDNA synthesis. This technique yields up to 10^5- fold linear amplification of high-fidelity aRNA from nanograms of total RNA and is applicable whether total or poly(A) RNA is used. These results define the operational parameters of RNA amplification approaches and expands the utilization of cDNA microarrays to experimental conditions in which starting material is the limiting factor. These include clinical specimens from fine needle aspirates or micro-dissection or experimental models studying embryonic tissue or small organisms.

Example 3: Target preparation for gene profiling (transcriptome) array

This example provides a method for the preparation of nucleic acid samples (targets) for applying on a gene profiling array.

First, aRNA is amplified. Total RNA is isolated from a biological sample, such as a fresh or preserved cell or tissue sample or an aliquot of cells grown in culture. By way of example, total RNA was isolated using a Qiagen midi kit (Cat. #75142) following the instructions provided by the manufacturer. Alternatively, Trizol extraction (Gibco BRL Cat. # 15596-026) could also be used (following the procedures provided by the manufacturer). The total RNA was then resuspended or eluted in DEPC water.

First strand cDNA synthesis was carried out as follows: In a PCR reaction tube, 0.001-3 µg total RNA was mixed in 9 µl DEPC H_2O with 1 µl (0.01-0.5 µg/µl) oligo $dT_{(15)}$-T7 primer (SEQ ID NO: 1) and heated to 70° C for three minutes, then cooled to room temperature. To this was then added the following reagents (which can be made into a "mastermix" for multiple samples):

4 µl 5 X First strand buffer (provided with Superscript II kit)
1 µl (0.1-0.5 µg/µl) TS (template switch) oligo primer (SEQ ID NO: 3)
2 µl 0.1M DTT
1 µl RNAsin (Promega Cat. # N2111)
2 µl 10mM dNTP (Pharmacia Cat. # 27-2035-02)
2 µl Superscript II polymerase (Gibco BRL Cat. # 18064-071)

The reaction was then incubated 42 °C for 90 minutes in a thermal cycler.

Second strand synthesis was carried out by adding the following reagents to each cDNA reaction tube:

106 µl of DEPC H_2O
15 µl Advantage PCR buffer
3 µl 10 mM dNTP
1µl of RNase H (2U/µl, Gibco BRL Cat# 18021-071)
3µl Advantage Polymerase (Clontech Cat# 8417-1)

The samples were then incubated at 37°C for five minutes to digest mRNA, 94°C for two minutes to denature, 65° C for one minute for specific priming, and 75° C for 30 minutes for extension of the second strand. The reaction was stopped by adding 7.5 µl 1M NaOH solution containing 2 mM EDTA and incubating at 65°C for 10 minutes to inactivate enzyme.

The double stranded (ds) cDNA was cleaned up as follows: A 1 µl aliquot of Linear Acrylamide (0.1 µg/µl, Ambion Cat. # 9520) was added to each sample. The sample was then extracted by adding 150 µl Phenol: Chloroform: Isoamyl alcohol (25:24:1) (Boehringer Mannhem Cat. #101001) to each ds cDNA tube and mixing well by pipetting. It is important not to be careful not to spill or contaminate the sample. The slurry solution was then transferred to Phase lock gel tube (5'-3' Inc. Cat. # p1-257178) and spun at 14,000 rpm for five minutes at room temperature. The

aqueous phase was transferred to RNase/DNase-free tube and 70 µl of 7.5M ammonium acetate (Sigma Cat# A2706) added, followed by 1 ml 100% ethanol. This tube was centrifuged at 14,000 rpm for 20 minutes at room temperature to pellet the nucleic acid. The resultant pellet was washed twice with 500 µl 100% ethanol and spun down at maximum speed for eight minutes. Finally, the ds cDNA pellet was air dried and resuspended in 70 µl DEPC H_2O.

Bio-6 Chromatograph columns (Bio-Rad Cat. # 732-6222) were prepared by washing the columns with 700 µl DEPC H_2O three times and spinning at 700 xg for two minutes at room temperature. (It may be important to shake the washed column well before draining to get rid of air bubbles - otherwise it drains very slowly.) When opening the column, any gel in the underside of the cap was aspirated off to prevent contamination. Also, the collection tubes provided with Bio-6 columns are not RNase-free; the samples should be collected in RNase-free tubes.

For each sample, 70 µl was loaded onto the center of the column and the column spun at 700x g for four minutes. The sample was then dried by Speedvac and resuspended in 8 µl DEPC water.

Using this double-stranded cDNA, *in vitro* transcription (IVT) was performed using an Ambion T7 Megascript Kit (Cat. #1334). For each sample, the following reaction mixture was made:

2 µl of each 75 mM NTP (A, G, C and UTP)
2 µl reaction buffer
2 µl enzyme mix (RNase inhibitor and T7 phage polymerase)
8 µl ds cDNA (produced as described herein)

The reactions were then incubated at 37° C for 6 hours to permit transcription.

The aRNA produced was then purified using TRIzol reagent (GibcoBRL, Cat. #15596). To each IVT reaction was added 1 ml of TRIzol solution, and the tubes were mixed well. 200 µl of chloroform was then added per 1 ml TRIzol solution, and the samples mixed by inverting for 15 seconds. They were then incubated at room temperature for 2-3 minutes, and centrifuged at 12,000g for 15 minutes at 4°C. The aqueous phase was then transferred to a new RNase free tube and 500 µl of isopropyl alcohol added per 1 ml TRIzol reagent to precipitate the nucleic acids. The samples were incubated at room temperature for 10 minutes and then centrifuged at 14,000 rpm for 15 minutes. The resultant pellet was washed two times with 1 ml 70% ethanol in DEPC-treated water, the pellet air dried and quickly resuspended in 20 µl DEPC-treated water. (Over-dried RNA is difficult to dissolve into water). RNA concentration can be checked and quality estimated by measuring OD_{260} and $OD_{260}/_{280}$ using standard techniques.

An RNAeasy mini kit also could be used to recover the aRNA (but the recovery of aRNA is lower compared with the TRIzol method.)

The aRNA was subjected to a second round of amplification, though this is not necessary in all embodiments. By way of example, aRNA (0.5-1 µg) produced as above was mixed in 9 µl

DEPC H_2O with 1 µl (2 µg/µl) random hexamer (*i.e.*, dN6) and heated to 70°C for three minutes, then cooled to room temperature. The following reagents were then added:

4 µl 5 X First strand buffer
1 µl (0.5 µg/µl) oligo dT-T7 primer
2 µl 0.1M DTT
1 µl RNAsin (Promega Cat. # N2111)
2 µl 10 mM dNTP (Pharmacia Cat. # 27-2035-02)
2 µl Superscript II (SS II) (Gibco BRL Cat. # 18064-071)

The samples were then incubate at 42° C for 90 minutes. The resultant single-stranded cDNA can then be subjected to second strand synthesis and cleanup similarly to that described above. In this example, the ds cDNA was then resuspended in 16 µl of DEPC treated water.

Second round *in vitro* transcription (IVT) proceeded using the following reaction mixture:

4 µl of each 75 mM NTP (A, G, C and UTP)
4 µl reaction buffer
4 µl enzyme mix (RNase inhibitor and T7 phage polymerase)
16 µl ds cDNA

Each reaction was incubated at 37° C for six hours, and the aRNA purified using TRIzol reagent, as described.

In order to prepare target nucleic acid to be printed on transcriptome (gene profiling) arrays, aRNA amplified from the second IVT was first converted into cDNA using the following reverse transcription reaction:

6 µg of aRNA (1 µg/µl)
2 µl of dN6 primer (8 µg/µl)
14 µl of DEPC treated water

Samples were heated to 70°C for three minutes and then put on ice. Then, the following reagents were added:

8 µl of 5X first strand buffer
4 µl of 10 mM dNTP
4 µl of 0.1M DTT
2 µl of RNAsin
3 µl of Superscript II

The samples were then incubated at 42°C for 90 minutes. The reactions were stopped by adding 5 µl of 0.5M EDTA with 10 µl of 1M NaOH and heating to 65°C for 10 minutes, which hydrolyzed the aRNA and inactivated the enzymes. The pH of the samples was neutralized by adding 25 µl of 1M Tris pH 7.5.

Target nucleic acids were purified (precipitated) as follows: To each sample was added 30 µl of ammonium acetate and 500 µl 100% ethanol, and the samples were mixed and incubated at -20° C for 15 minutes. Samples were centrifuged at 13,000 rpm at 4°C for 20 minutes, and the resultant

pellet washed twice with 500 µl of 70% ethanol. The pellet was then completely dried using a Speedvac, and the purified cDNA resuspended in 12.5 µl of 3X SSC; in some embodiments, to get a stronger signal the cDNA is resuspended in a smaller volume. Resuspended cDNA can be stored at -20°C.

Internal control genes printed onto the array can be any known "house keeping" gene, in other words a gene expected not to be affected by the test situation (*e.g.*, not altered in the cancer being tested). By way of example only, β-actin was used as an internal control gene. A specific 5' primer and modified 3' specific primer (with T7 promoter region) were designed (using information available in public databases) to flank a 400 base pair sequence close to the poly A tail. After PCR amplification, a 400 bp double strand β-actin with the T7 promoter attached to the 5' end was produced.

The PCR product was then cleaned up essentially as provided above for ds cDNA cleanup. For each sample, 1 µg of PCR product was used as a template for *in vitro* transcription to generate sense β-actin RNA, and then converted into cDNA (as described), and two-fold and 10-fold serial β-actin dilutions (from 6 µg to 60 pg in 12.5 µl of 3X SSC) were used for printing control samples.

Target solutions were transferred to a 384 well U bottom micro-plate, and printed to slides using a GeneMachine robotic printer, as described in Example 1 except that a 4 pen was used instead of a 32 pen.

A list of samples used to make a transcriptome array is shown in Table 1. **Column labels:** Sample ID – reference name unique for each different sample; cell type – type of cell or cell line from which nucleic acid originated; sample name – additional descriptive information regarding individual samples; aRNA 2IVT – quantification of aRNA produced after second round of *in vitro* transcription (µg/ml); volume – volume of DEPC H_20 used to resuspend aRNA.

Example 4: Probe preparation for hybridization

Gene (Pmel-17 and RhoC) -specific PCR primer sets and house keeping gene (β actin) – specific PCR primer sets were designed using the Primer 3 program, developed by Whitehead Institute, based on full-length cDNA sequence of each of these molecules from data base. The size of each amplicon was selected to be about 350-450 bp, with a 3' bias. PCR products with incorporated Cy3 (for specific gene) and Cy5 (for house keeping gene) can be applied for hybridization after purification and denaturalization. Methods for integrating Fluorolink Cy5 or Cy3-dUTP or other fluorescent molecules into nucleic acids are well known.

Modification of the 5' primer by attachment of a T7 promoter region can also be used for probe preparation. PCR products with a T7 promoter extension region can then be used as template in *in vitro* transcription (IVT) to generate sense RNA that will be converted into cDNA in the presence of Cy5 or Cy3 labeled nucleotides, thereby providing incorporation of Cy3 or Cy5.

Fluorescent labeled ss cDNAs were used in the hybridization example presented herein. Labeled ss cDNAs were prepared using the following reaction mixture:

4 µl First strand buffer
1 µl dN6 primer (8 µg/µl; Boehringer Mannheim Cat. # 1034731, re-suspended in 250 µl DEPC H_2O)
2 µl 10X low T - dNTP (5mM A, C and GTP, 2 mM dTTP)
2 µl Cy-dUTP (1 mM Cy3 or Cy5)
2 µl 0.1 M DTT
2 µl RNasin
3 µg amplified aRNA in 16 µl DEPC H_2O

Reactions were mixed well and heated to 65°C for five minutes, then cooled to 42°C. To each reaction was added 1 µl Superscript II polymerase. The samples were then incubate for 30 minutes at 42°C, another 1 µl polymerase added and the incubation continued for an additional 40 minutes at 42°C. To stop the reaction, 2.5 µl 500 mM EDTA was added and the samples heated to 65°C for one minute. Then 5 µl 1M NaOH was added and the samples incubated at 65°C for 15 minutes to hydrolyze the RNA. Tris buffer (12.5 µl of 1M) was added immediately to neutralize the pH, and the volume raised to 70 µl by adding 35 µl of 1x TE.

Nucleic acid probes were cleaned up using Bio-6 columns, which were prepared and run essentially as described above. Flow through was collected and 200 µl 1 x TE added to each. The probe preparation was then concentrated to a volume of ~20 µl using microcon YM-30 column (Millipore Cat. #42410).

Example 5: Hybridization

Cy3 and Cy5 labeled probe were combined (1:1 ratio) and concentrated to 16 µl using a speed vacuum. To each sample was added:

1 µl of 50x Denhardt's blocking solution (Sigma Cat. # 2532)
1 µl poly dA (8 µg/µl Pharmacia Cat. # 27-7988-01)
1 µl yeast tRNA (4 mg/ml Sigma Cat. # R8759)
1 µl Human Cot I DNA (10 mg/ml Gibco BRL Cat#15279-011)
2.6 µl 20X SSC.

The samples were then heated for two minutes at 99°C, and 0.6 µl of 10% SDS added. Samples were then cooled to room temperature. Prepared probe mixture was applied to an array slide, a cover slip added, and the slide placed in a humidified hybridization chamber. The samples were allowed to hybridize at 65°C overnight.

The slides were washed using the following washing protocol:

2x SSC + 0.1% SDS to get rid of the cover slide
1x SSC for one minute
0.2 x SSC for one minute
0.05x SSC for 10 second

Washed slides were centrifuged gently at 80-100x g for three minutes to remove excess liquid. (Slide can be put in slide rack on microplate carriers or in 50 ml conical tube and centrifuged in swinging-

bucket rotor.) The slides were then scanned for fluorescent signals using a commercially available scanner GenePix 4000B and GenePix Pro3 software, from Axon Instruments, Inc:.

Having illustrated and described the principles of gene profiling microarrays, wherein a plurality of mixtures of nucleic acids from a collection of different specimens are arrayed together on an array, and the use of such arrays for analysis of gene expression, and various methods for production of the mixtures of nucleic acids arrayed and probes used to assay them, it will be apparent that the invention can be modified in arrangement and detail without departing from such principles. In view of the many possible embodiments to which the principles of this invention may be applied, it should be recognized that the illustrated embodiments are only examples of the invention and should not be taken as a limitation on the scope of the invention. Rather, the scope of the invention is in accord with the following claims. We therefore claim as our invention all that comes within the scope and spirit of these claims.

Table 1.

Sample ID	cell type	sample name	aRNA 2IVT	vol
CTL 001	CTL	Reed alpha flu CTL C30 (3/24/96)	2052.50	40
EBV 001	B cell	583 EBV	1770.50	40
EBV 002	B cell	888 EBV (7/15/96)	1223.80	40
FRTN 001	melanoma tumor digest	Genc TC 1703-3 (FRTN)	2296.60	40
L 001	lymphoma	MLI	235.19	30
M 001	melanoma cell line	M.Goldman	2749.20	30
M 002	melanoma cell line	888 Mel	3168.00	30
M 014	melanoma cell line	WC 013 Mel	1590.40	40
PBMC 001	PBMC	Anderson (12/18/96)	3740.10	40
PBMC 002	PBMC	Avery (6/8/98)	3975.30	40
PBMC 003	PBMC	Barnes (10 /8/97)	3302.10	40
F19112	kerolinocyte	F19112	4872.20	30
FF2440	fibroblast	FF2440	7023.80	30
Hmvec	endothelial	Hmvec	4144.00	30
FNA12	melanoma	Post 2M and IL-2	3843.90	40
FNA94	melanoma	Post 2M and IL-2	1915.00	40
FNA158	melanoma	Post A1 and IL-2	3264.30	40
FNA160	melanoma	Post A1 and IL-2	2828.00	40
FNA159	melanoma	Post A1 and IL-2	2145.20	40
FNA104	melanoma	Post A3 and IL-2	2042.80	40
FNA33	melanoma	Post A31 and IL-2	3047.20	40
FNA79	melanoma	Post 2M and IL-2	3190.30	40
FNA99	melanoma	Post two doses GP Mart and IL-2	3584.70	40
FNA129	melanoma	Rsctd spec/ post ES pep+sq IL-2	1242.30	40
FNA44	melanoma	Post ES peptide alone	3126.80	40
FNA165	melanoma	Post ES peptide, MART	1042.10	40
FNA126	melanoma	post gp/MART peptide alone	1226.60	40
FNA91	melanoma	GP100 q week	1248.10	40
FNA92	melanoma	GP100 q week	2865.80	40

Sample ID	cell type	sample name	aRNA 2IVT	vol
FNA71	melanoma	Post 2M peptide vaccines	2565.40	40
FNA73	melanoma	Post 2M peptide vaccines	3217.20	40
FNA74	melanoma	Post 2M peptide vaccines	4361.80	40
FNA75	melanoma	Post 2M peptide vaccines	2543.00	40
FNA77	melanoma	Post 2M peptide vaccines	3724.10	40
FNA76	melanoma	Post 2M peptide vaccines	1448.60	40
FNA78	melanoma	Post 2M peptide vaccines	647.80	40
FNA89	melanoma	Pre tx	2293.00	40
FNA98	melanoma	Pre tx	3341.90	40
FNA108	melanoma	Pre tx	1277.10	40
FNA109	melanoma	Pre tx	923.40	40
FNA176	melanoma	Pre tx	3729.80	40
FNA150	melanoma	Pre tx	1542.50	40
FNA171	melanoma	Pre tx	8097.70	40
FNA149	melanoma	Pre tx	3355.40	40
FNA31	melanoma	Pre tx	1591.20	40
FNA100	melanoma	Pre tx	6034.30	40
control	Beta actin	6ug/12.5ul		
control	Beta actin	0.6ug/12.5ul		
control	Beta actin	0.06ug/12.5ul		
control	Beta actin	0.006ug/12.5ul		
control	Beta actin	0.0006ug/12.5ul		
control	Beta actin	0.00006ug/12.5ul		
control	Beta actin	0.6ug/12.5ul		
control	Beta actin	0.3ug/12.5ul		

(Table 1 continued)

We claim:

1. A gene expression assay method comprising:

providing an array of nucleic acid mixtures at addressable locations on a substrate, wherein the nucleic acid mixtures comprise nucleic acid molecules in quantities that are substantially proportional to quantities of the nucleic acid molecules in a specimen from which the nucleic acid molecules are obtained; and

exposing the array to at least one probe for detecting one or more nucleic acid molecules on the array under conditions sufficient to produce binding of the probe to the one or more nucleic acid molecules.

2. The method of claim 1, further comprising detecting binding of the probe.

3. An assay method to determine relative expression of a DNA sequence in a plurality of biological specimens, the assay method comprising:

(a) providing a labeled probe;

(b) contacting the labeled probe with an array of mixtures of nucleic acid molecules arrayed on a surface of a solid support, under conditions sufficient to produce binding, wherein each mixture of nucleic acid molecules proportionately reflects expression levels of RNA molecules from a specimen from which the nucleic acid molecules are obtained;

(c) separating unbound labeled probe from the array; and

(d) detecting probe binding on the array.

4. The method of claim 2 or 3, wherein detecting comprises quantitatively detecting binding to yield an amount of probe binding which correlates with the expression levels of RNA molecules.

5. The method of claim 4, further comprising correlating the amount of probe binding with a level of gene expression in the specimen.

6. The method of claim 1 or 3, wherein the nucleic acid mixtures are stably associated with a surface of the substrate.

7. The method of claim 1 or 3, wherein the specimens are selected from the group consisting of cells or tissues.

8. The method of claim 7, wherein the cells comprise animal, microbial or plant cells.

9. The method of claim 8, wherein the animal cells comprise human cells.

10. The method of claim 1 or 3, wherein the probe is a nucleic acid molecule having specific complementarity to a target RNA molecule.

11. The method of claim 10, wherein the probe is a single-stranded nucleic acid.

12. The method of claim 1 or 3, wherein each mixture substantially proportionately reflects the expression level of substantially all expressed mRNA molecules of that specimen.

13. The method of claim 1 or 3, wherein the mixtures of nucleic acids comprise mixtures of amplified nucleic acid molecules.

14. The method of claim 13, wherein the nucleic acid molecules are amplified prior to detecting binding of the probe.

15. The method of claim 13, wherein the nucleic acid molecules are amplified prior to placement on the array.

16. The method of claim 13, wherein the mixtures of amplified nucleic acid molecules are amplified by a method comprising:

isolating an RNA sample from a specimen;

obtaining one or more RNA templates from a portion of the RNA sample;

hybridizing the one or more templates with a first primer to form a primed template, wherein the first primer comprises an antisense sequence of an RNA polymerase promoter;

synthesizing first strand cDNA from the primed template;

hybridizing the first strand cDNA with a second primer to form a switched template, wherein the second primer has a 5' end and a 3' end and comprises a string of dG residues at the 3' end;

synthesizing second strand cDNA from the switched template to generate full-length double stranded cDNA;

transcribing aRNA from the full-length double stranded cDNA; and

reverse transcribing amplified cDNA from the transcribed aRNA.

17. The method of claim 16, wherein the mixture of amplified nucleic acid molecules substantially proportionately reflects the expression level of substantially all expressed mRNA molecules of a specimen from which the nucleic acid molecules are obtained.

18. The method of claim 2 or 3, further comprising substantially removing unbound probe molecule prior to detecting the binding.

19. The method of claim 1 or 3, wherein the probe comprises a detectable tag.

20. The method of claim 19, wherein the detectable tag comprises a fluorophore, a radioactive isotope, a ligand, a chemiluminescent agent, a metal sol, a metal colloid, or an enzyme.

21. The method of claim 20, wherein the tag comprises a fluorophore.

22. The method of claim 1 or 3, further comprising applying two or more differently labeled probes simultaneously or sequentially and reading the hybridization pattern of both labels.

23. The method of claim 22, wherein the differently labeled probes are labeled with fluorophores of different colors.

24. The method of claim 22, wherein one of the differently labeled probes is a control probe.

25. The method of claim 24, wherein the control probe corresponds to a housekeeping gene.

26. The method of claim 6, wherein the mixtures of nucleic acid molecules are associated with the substrate at discrete addresses.

27. The method of claim 26, wherein at least half of the mixtures of nucleic acid molecules are from different specimens.

28. The method of claim 2 or 3, wherein the binding detected is a binding pattern.

29. The method of claim 2 or 3, wherein detecting comprises automated detection.

30. A gene profiling array, comprising

a plurality of mixtures of nucleic acid molecules immobilized on a solid support in an addressable pattern, and

wherein each mixture proportionately reflects expression levels of mRNA molecules in a specimen from which the nucleic acid molecules are obtained.

31. The array of claim 30, wherein different mixtures of nucleic acid molecules are derived from a plurality of different specimens.

32. The array of claim 30, wherein the addressable pattern comprises mixtures of nucleic acid molecules in discrete spots, the spots arranged in rows and columns.

33. The array of claim 30, wherein the addressable pattern is arranged in a computer readable format.

34. The array of claim 30, comprising at least 10 different mixtures of nucleic acid molecules.

35. The array of claim 30, comprising at least 30 different mixtures of nucleic acid molecules.

36. The array of claim 30, comprising at least 100 different mixtures of nucleic acid molecules.

37. The array of claim 30, wherein the array comprises a microarray.

38. The array of claim 37, wherein the mixtures of nucleic acid molecules are in spots, and the spots have a maximum dimension of about 1 millimeter.

39. The array of claim 30, wherein the solid support comprises glass, nitrocellulose, polyvinylidene fluoride, nylon, fiber, or combinations thereof.

40. The array of claim 30, wherein the specimens are selected from the group consisting of cells and tissues.

41. The array of claim 30, wherein the cells comprise animal, plant or microbial cells.

42. The array of claim 30, wherein the mixtures of nucleic acid molecules are amplified prior to being immobilized on the solid support.

43. The array of claim 42, wherein amplifying the mixtures of nucleic acid molecules prior to immobilizing them on the solid support comprises:

isolating an RNA sample from a specimen;

obtaining one or more RNA templates from a portion of the RNA sample;

hybridizing the one or more templates with a first primer to form a primed template, wherein the first primer comprises an antisense sequence of an RNA polymerase promoter;

synthesizing first strand cDNA from the primed template;

hybridizing the first strand cDNA with a second primer to form a switched template, wherein the second primer has a 5' end and a 3' end and comprises a string of dG residues at the 3' end;

synthesizing second strand cDNA from the switched template to generate full-length double stranded cDNA;

transcribing aRNA from the full-length double stranded cDNA; and

reverse transcribing amplified cDNA from the transcribed aRNA.

44. A kit for determining relative expression of a DNA sequence in a plurality of biological specimens, comprising

the gene profiling array of claim 30; and

instructions for using the array.

45. The kit of claim 44, further comprising a probe representing the DNA sequence.

46. The kit of claim 44, wherein the instructions include directions for exposing at least one probe molecule to the gene profiling array to detect and/or quantify gene expression.

47. The kit of claim 44, wherein the gene profiling array comprises a microarray.

48. The kit of claim 44, further comprising a buffer.

49. The kit of claim 44, further comprising a probe.

50. The kit of claim 49, wherein the probe comprises a label.

51. The kit of claim 44, further comprising a probe standard.

52. The kit of claim 51, wherein the probe standard comprises a label.

53. The kit of claim 44, wherein the specimens are selected from the group consisting of cells and tissues.

54. The kit of claim 53, wherein the cells comprise animal, plant or microbial cells.

55. An assay method for analyzing a plurality of gene expression profiles, comprising:

(a) providing the array of claim 30;

(b) exposing the array to a first probe that may hybridize to the nucleic acid molecules of the array to identify those nucleic acid molecules to which the first probe hybridizes;

(c) detecting a first hybridization pattern of the first probe;

(d) repeating (b) through (c) with a second probe to identify samples to which the second probe hybridizes.

56. The method of claim 55, further comprising stripping hybridized first probe from the array prior to exposing the array to the second probe.

57. The method of claim 55, wherein either the first probe or the second probe is a control probe.

58. A method of producing a mixture of mRNA-derived nucleic acid molecules, comprising:

isolating an RNA sample from a specimen;

obtaining one or more RNA templates from a portion of the RNA sample;

hybridizing the one or more templates with a first primer to form a primed template, wherein the first primer comprises an antisense sequence of an RNA polymerase promoter;

synthesizing first strand cDNA from the primed template;

hybridizing the first strand cDNA with a second primer to form a switched template, wherein the second primer has a 5' end and a 3' end and comprises a string of dG residues at the 3' end;

synthesizing second strand cDNA from the switched template to generate full-length double stranded cDNA;

transcribing aRNA from the full-length double stranded cDNA; and

reverse transcribing amplified cDNA from the transcribed aRNA.

59. A gene expression assay method comprising:

providing an array of nucleic acid mixtures at addressable locations on a substrate; and

exposing the array to at least one probe for detecting one or more nucleic acid molecules on the array under conditions sufficient to produce binding of the probe to the one or more nucleic acid molecules.

60. The method of claim 59, further comprising detecting binding of the probe.

61. The method of claim 59, wherein at least half of the mixtures of nucleic acid molecules are from different specimens.

62. The method of claim 60, wherein the binding detected is a binding pattern.

63. The method of claim 60, wherein detecting comprises automated detection.

64. The method of claim 1, 3, or 59, wherein the array comprises a microarray.

65. The method of claim 59, wherein at least one mixture of nucleic acids is derived from a specimen consisting of not more than 10 cells.

66. The method of claim 65, wherein the specimen consists of not more than one cell.

67. The method of claim 59, wherein at least one nucleic acid mixture is derived from a source RNA sample extracted from a source specimen, and wherein the source RNA sample consists of no more than about 1 µg of total RNA.

68. The method of claim 67, wherein the source RNA sample consists of no more than about 0.75 µg of total RNA.

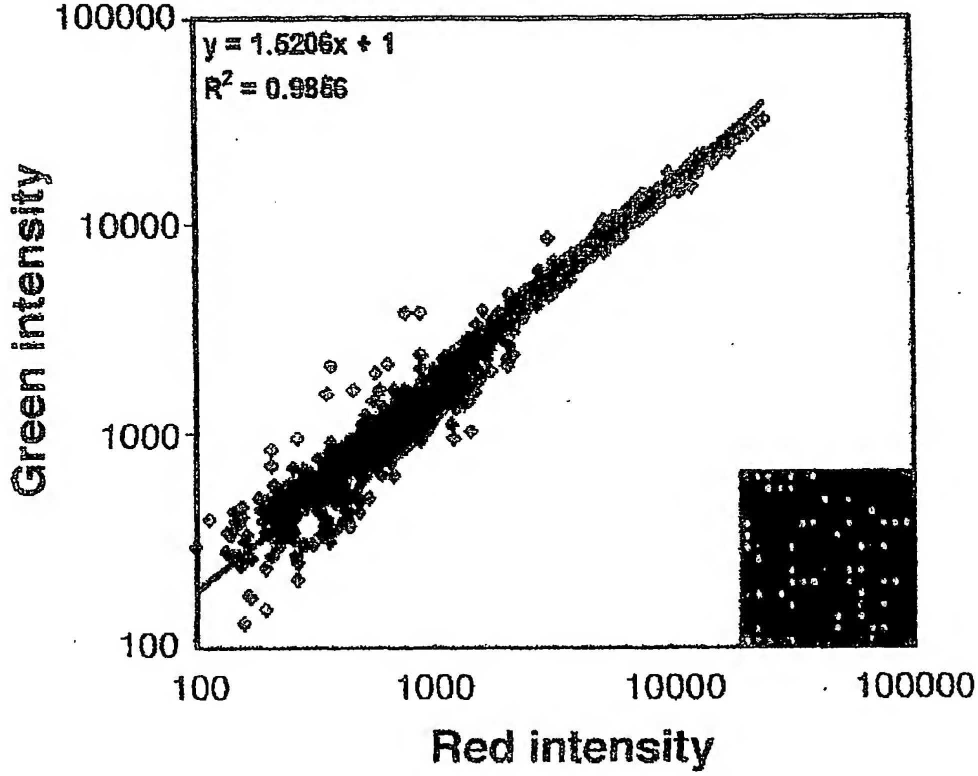
69. The method of claim 67, wherein the source RNA sample consists of no more than about 0.5 µg of total RNA.

70. The method of claim 67, wherein the source RNA sample consists of no more than about 0.3 µg of total RNA.

71. The method of claim 59, wherein the array comprises at least 100 different mixtures of nucleic acid molecules.

72. An array used in the method of claim 59.

Figure 1


y = 1.5206x + 1
R^2 = 0.9866
100000
10000
1000
100
Green intensity
100
1000
10000
100000
Red intensity

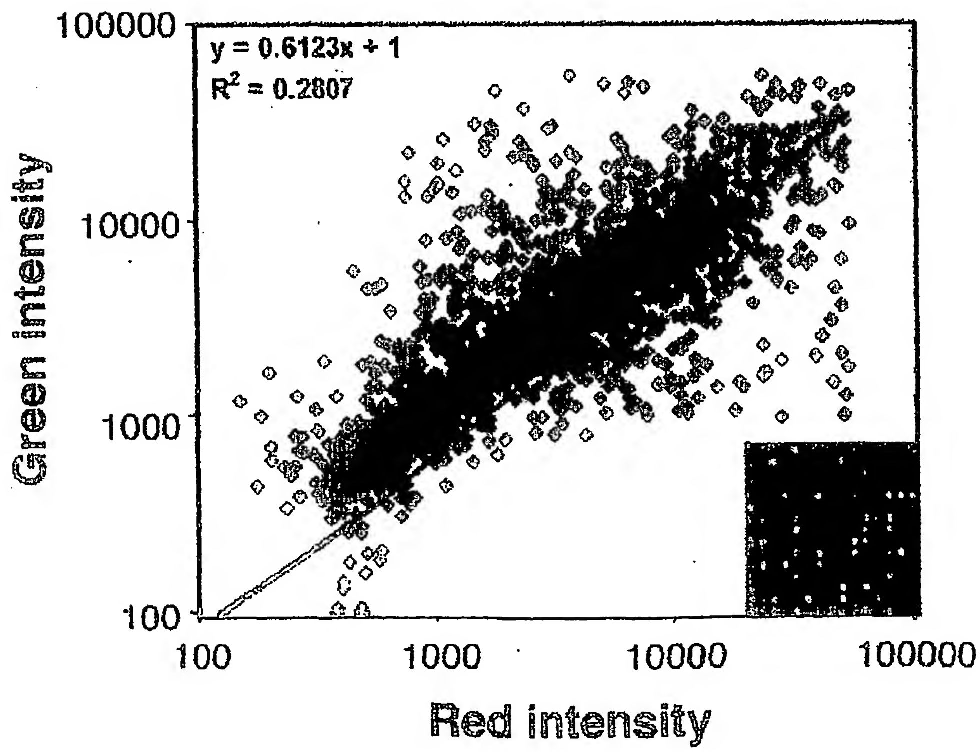
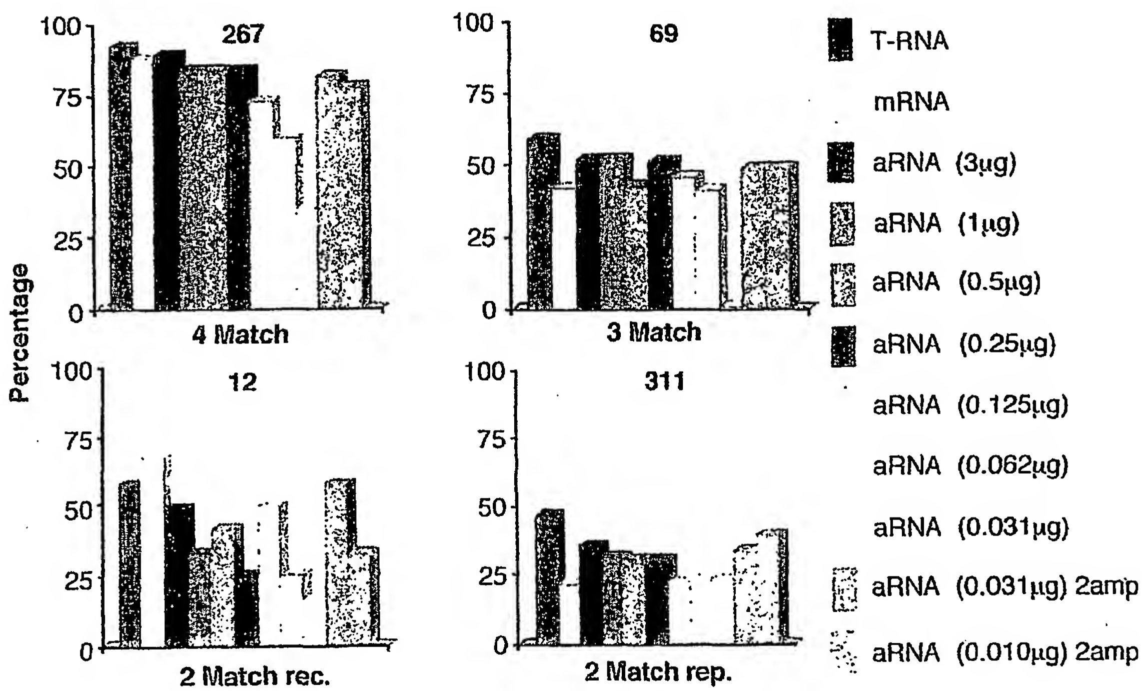

y = 0.6123x + 1
R^2 = 0.2807
100000
10000
1000
100
Green intensity
100
1000
10000
100000
Red intensity

Figure 2a



267
69
12
311
Percentage
100
75
50
25
0
4 Match
3 Match
2 Match rec.
2 Match rep.
T-RNA
mRNA
aRNA (3µg)
aRNA (1µg)
aRNA (0.5µg)
aRNA (0.25µg)
aRNA (0.125µg)
aRNA (0.062µg)
aRNA (0.031µg)
aRNA (0.031µg) 2amp
aRNA (0.010µg) 2amp

Figure 2b

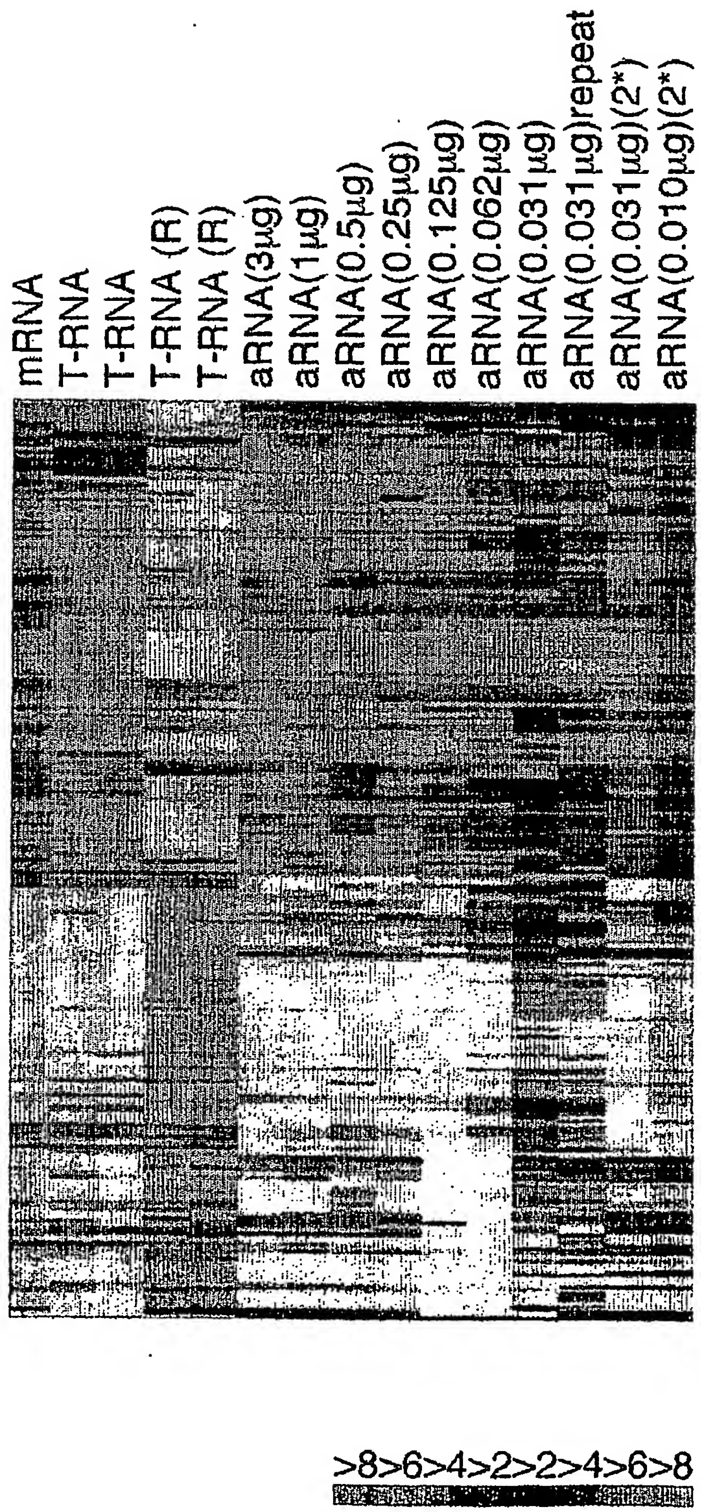

mRNA
T-RNA
T-RNA
T-RNA (R)
T-RNA (R)
aRNA(3μg)
aRNA(1μg)
aRNA(0.5μg)
aRNA(0.25μg)
aRNA(0.125μg)
aRNA(0.062μg)
aRNA(0.031μg)
aRNA(0.031μg)repeat
aRNA(0.031μg)(2*)
aRNA(0.010μg)(2*)
>8>6>4>2>2>4>6>8

Figure 3a

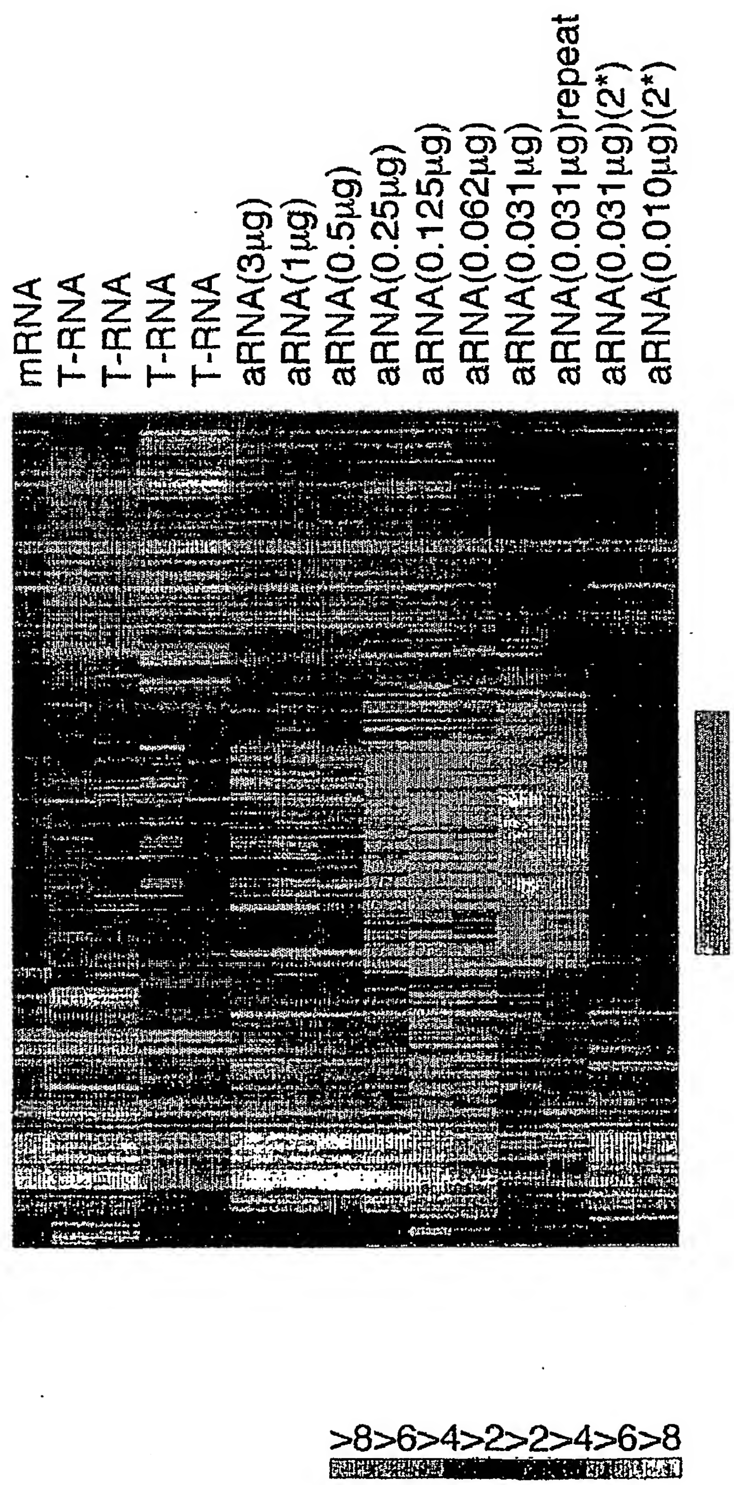

mRNA
T-RNA
T-RNA
T-RNA
T-RNA
aRNA(3µg)
aRNA(1µg)
aRNA(0.5µg)
aRNA(0.25µg)
aRNA(0.125µg)
aRNA(0.062µg)
aRNA(0.031µg)
aRNA(0.031µg)repeat
aRNA(0.031µg)(2*)
aRNA(0.010µg)(2*)
>8>6>4>2>2>4>6>8

Figure 3b

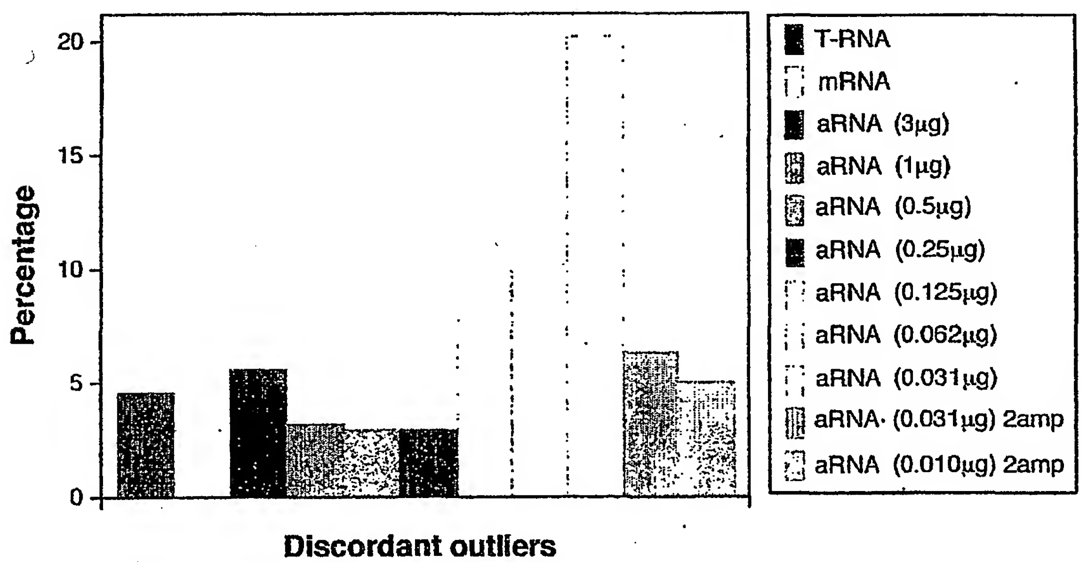

20
15
10
5
0
Percentage
Discordant outliers
T-RNA
mRNA
aRNA (3µg)
aRNA (1µg)
aRNA (0.5µg)
aRNA (0.25µg)
aRNA (0.125µg)
aRNA (0.062µg)
aRNA (0.031µg)
aRNA· (0.031µg) 2amp
aRNA (0.010µg) 2amp

Figure 4

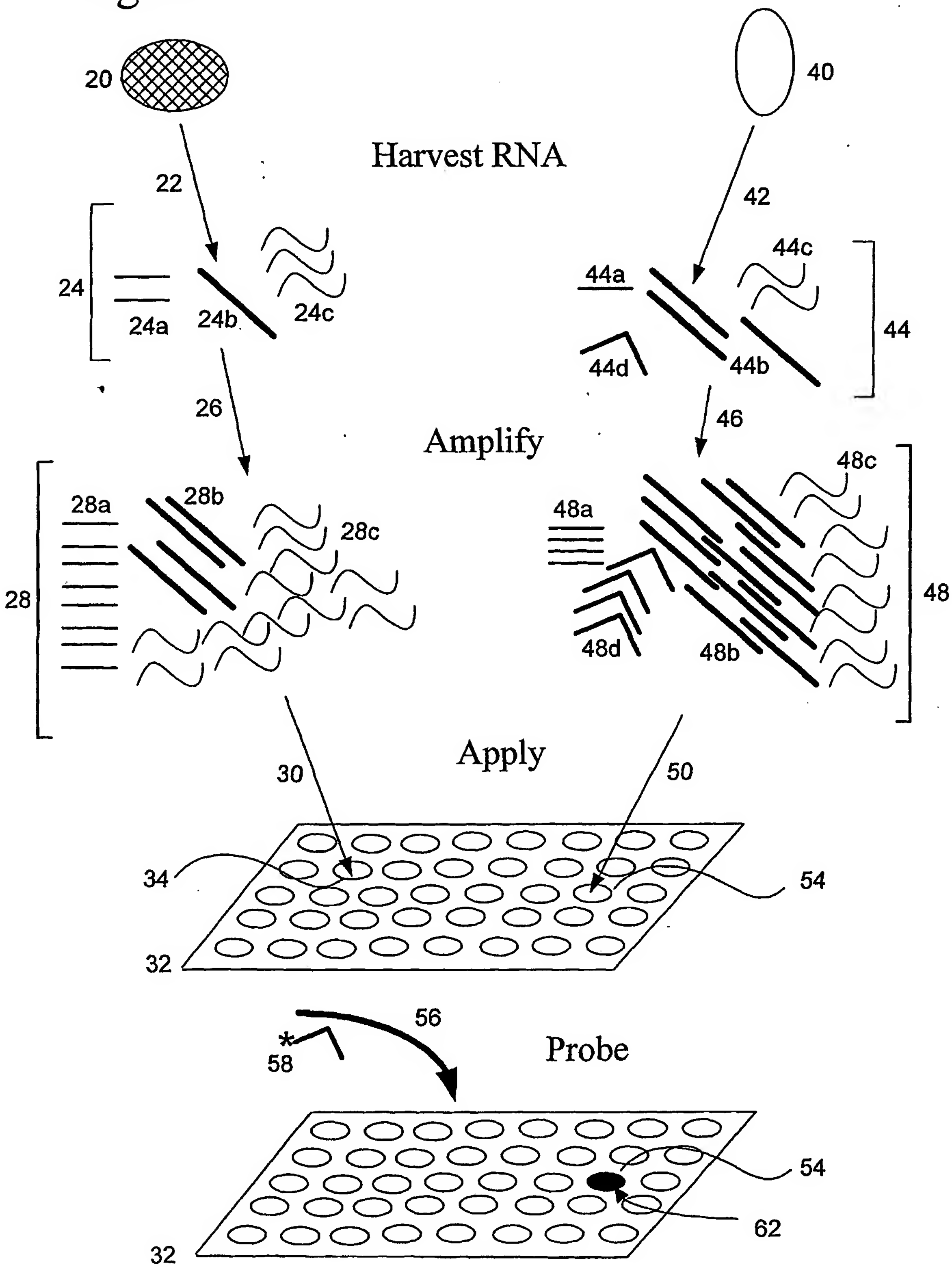

20
40
Harvest RNA
22
42
24
24a
24b
24c
44a
44c
44
44d
44b
26
46
Amplify
28a
28b
28c
28
48a
48c
48
48d
48b
Apply
30
50
34
54
32
56
58
Probe
54
62
32

Figure 5

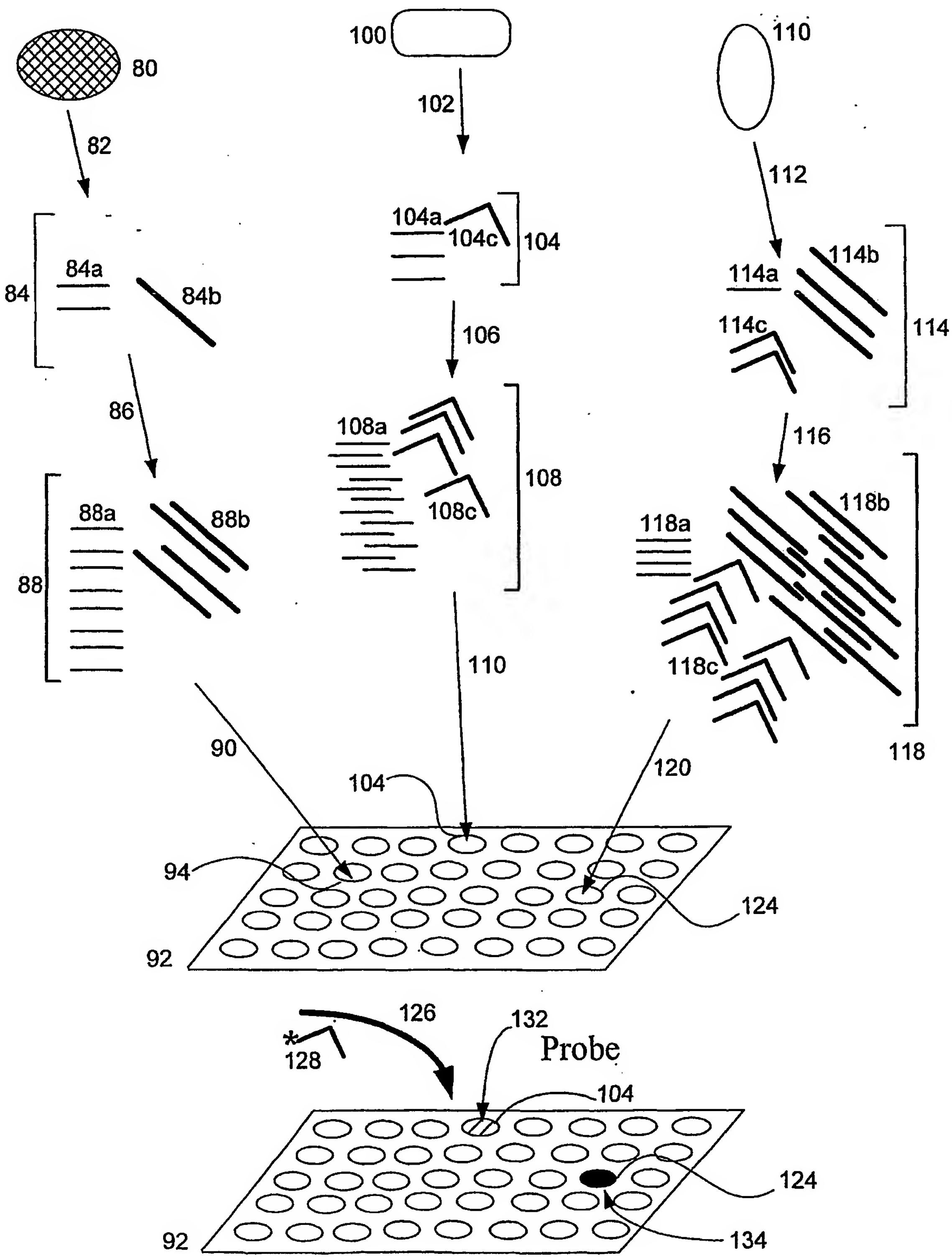
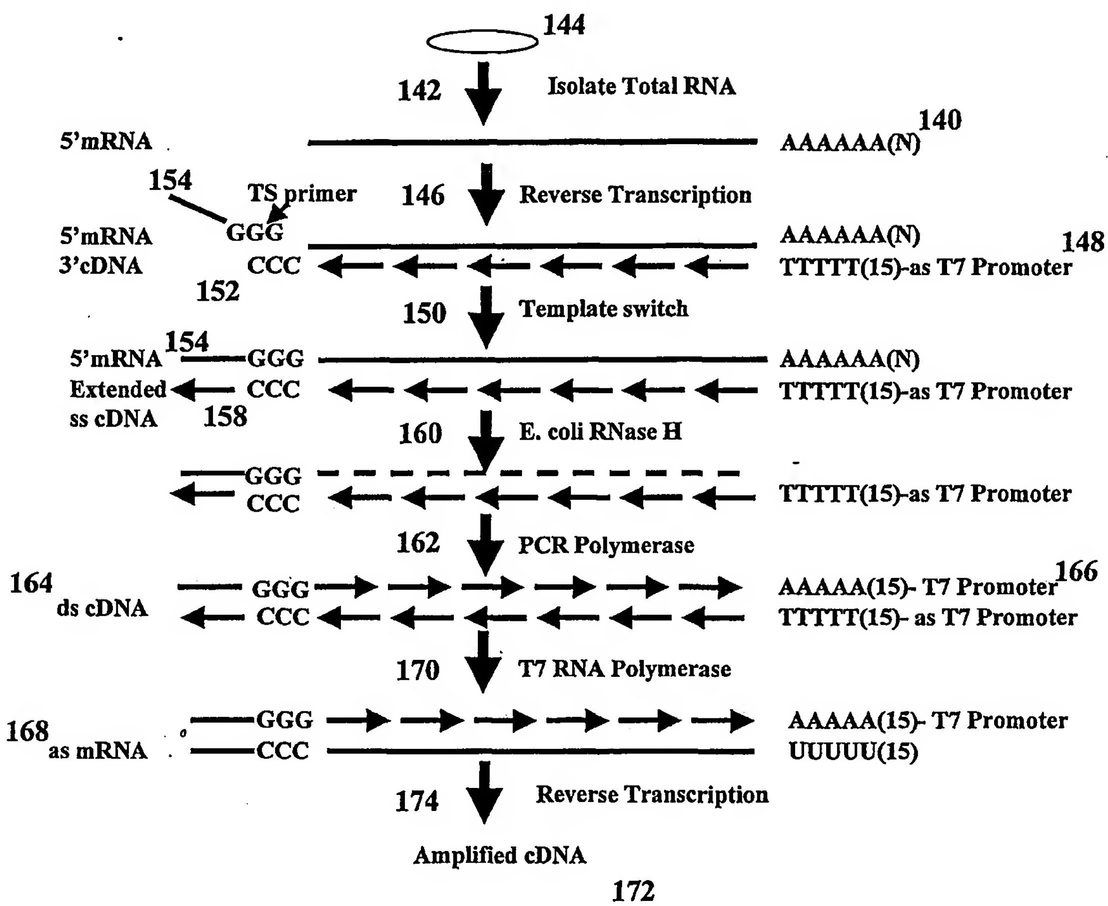

80
82
84
84a
84b
86
88
88a
88b
90
100
102
104a
104c
104
106
108a
108c
108
110
110
112
114a
114b
114c
114
116
118a
118b
118c
118
120
104
94
124
92
126
*
128
132
Probe
104
124
92
134

Figure 6


144
142
Isolate Total RNA
5'mRNA
140
AAAAAA(N)
154
TS primer
146
Reverse Transcription
5'mRNA
GGG
AAAAAA(N)
148
3'cDNA
CCC
TTTTT(15)-as T7 Promoter
152
150
Template switch
5'mRNA
154
GGG
AAAAAA(N)
Extended
CCC
TTTTT(15)-as T7 Promoter
ss cDNA
158
160
E. coli RNase H
GGG
CCC
TTTTT(15)-as T7 Promoter
162
PCR Polymerase
164
ds cDNA
GGG
AAAAA(15)- T7 Promoter
166
CCC
TTTTT(15)- as T7 Promoter
170
T7 RNA Polymerase
168
as mRNA
GGG
AAAAA(15)- T7 Promoter
CCC
UUUUU(15)
174
Reverse Transcription
Amplified cDNA
172

SEQUENCE LISTING

<110> THE GOVERNMENT OF THE UNITED STATES OF AMERICA, AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF HEALTH & HUMAN SERVICES, THE NATIONAL INSTITUTES OF HEALTH
Wang, Ena
Marincola, Francesco
Lance, Miller

<120> GENE PROFILING ARRAYS

<130> 4239-58579

<150> 60/192,700
<151> 2000-03-28

<160> 3

<170> PatentIn version 3.0

<210> 1
<211> 55
<212> DNA
<213> synthetic oligonucleotide

<400> 1
aaacgacggc cagtgaattg taatacgact cactataggc gctttttttt ttttt 55

<210> 2
<211> 42
<212> DNA
<213> synthetic oligonucleotide

<400> 2
aaacgacggc cagtgaattg taatacgact cactataggc gc 42

<210> 3
<211> 30
<212> DNA
<213> synthetic oligonucleotide

<400> 3
aagcagtggt aacaacgcag agtacgcggg 30